



Essays in Financial and Housing Economics

Citation

McQuade, Timothy. 2013. Essays in Financial and Housing Economics. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12362595>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Essays in Financial and Housing Economics

A dissertation presented
by
Timothy J. McQuade
to
The Department of Economics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Economics

Harvard University
Cambridge, Massachusetts

April 2013

© 2013 Timothy J. McQuade

All rights reserved.

Essays in Financial and Housing Economics

Abstract

This dissertation presents four essays. The first chapter builds a real-options, term structure model of the firm incorporating stochastic volatility and endogenous default to shed new light on the value premium, financial distress, momentum, and credit spread puzzles. The paper uses recently developed methodologies based on asymptotic expansions to solve the model. The second chapter, coauthored with Adam Guren, presents a model that shows how foreclosures can exacerbate a housing bust and delay the housing market's recovery. Quantitatively, the model successfully explains aggregate and retail price declines, the foreclosure share of volume, and the number of foreclosures both nationwide and across MSAs. The third and fourth chapters, coauthored with Stephen W. Salant and Jason Winfree, discuss the economics of untraceable experience goods in a variety of settings. The third chapter drops the "small country" assumption in the trade literature on collective reputation and shows how large exporters like China can address severe problems assuring the quality of its exports. The fourth chapter demonstrates how regulations in the formal sector of developing countries can lead to a quality gap between formal and informal sector goods. It moreover investigates how changes in regulation affect quality, price, aggregate production, and the number of firms in each sector.

Contents

1	Stochastic Volatility and Asset Pricing Puzzles	1
1.1	Introduction	1
1.2	Literature Review	5
1.3	Structural Model of the Firm	7
1.3.1	Firm Dynamics - Physical Measure	7
1.3.2	Firm Dynamics - Risk Neutral Measure	9
1.3.3	Capital Structure	11
1.3.4	Equity Valuation	13
1.3.5	Debt Valuation	15
1.4	Methodology	16
1.4.1	Time Scales in Volatility Modeling	17
1.4.2	Asymptotic Approximation	19
1.4.3	Cumulative Default Probabilities	24
1.4.4	Debt Valuation under Exogenous Default	25
1.5	Equity Valuation	26
1.5.1	Financial Distress and Momentum	27
1.5.2	Book-to-Market Effects	28
1.6	Debt Valuation	30
1.6.1	Calibration	30
1.6.2	Credit Spreads on Intermediate/Long Maturity Debt	33
1.6.3	Intuition	36
1.7	Short Maturity Debt and Extensions	39

1.7.1	Multiscale Stochastic Volatility	41
1.7.2	Rare Disasters	45
1.8	Conclusion	47
2	How Do Foreclosures Exacerbate Housing Downturns?	49
2.1	Introduction	49
2.2	Empirical Facts	53
2.3	Housing Market Model	58
2.3.1	Setup	58
2.3.2	Numerical Methods	66
2.4	Basic Model Results and Mechanisms	67
2.4.1	Market Tightness, Choosey Buyer, and Compositional Effects	67
2.4.2	Qualitative Results	68
2.4.3	Isolating the Role of Each Effect	71
2.5	An Extended Model of Default	71
2.5.1	Default in the Extended Model	73
2.5.2	Starting the Downturn	75
2.6	Quantitative Analysis: Calibration and Amplification	76
2.6.1	Calibration	76
2.6.2	Strength of Amplification Channel and Comparative Statics	79
2.7	Cross-MSA Quantitative Analysis	81
2.8	Welfare and Policy Implications	87
2.8.1	Welfare	87
2.8.2	Delaying Foreclosure	88
2.8.3	Interest Rate or Payment Reductions	90
2.8.4	Principal Reduction	91
2.8.5	Other Policies	91
2.9	Conclusion	92
3	Markets with Untraceable Goods of Unknown Quality: Beyond the Small-Country Case	95
3.1	Introduction	95
3.2	Model of International Trade	97
3.3	Potential Remedies	102
3.3.1	Multilateral Remedies	102

3.3.2	Unilateral Remedies	103
3.4	Conclusion	104
4	Regulating Experience Goods in Developing Countries	106
4.1	Introduction	106
4.2	Short and Long-run Equilibrium in the Absence of Regulation	109
4.3	Consequences of Two Policies	113
4.4	Conclusion	117
	Bibliography	119
A	Supplement to Chapter 1	130
A.1	Derivations and Proofs	130
A.1.1	Proof of Theorem 1.3.1	130
A.1.2	Derivation of Equation (1.4.13)	133
A.1.3	Proof of Theorem 1.4.1	134
A.2	Contingent Claims of Growth Firms	138
A.2.1	Value of Assets in Place	138
A.2.2	Equity Valuation of Growth Firms	138
A.2.3	Debt Valuation of Growth Firms	139
A.2.4	Principal Order Terms	141
A.2.5	First-Order Correction Terms	142
A.3	Volatility Time Scales	142
B	Supplement to Chapter 2	146
B.1	Derivations and Proofs	146
B.1.1	Steady State	146
B.1.2	Dynamics	150
B.1.3	Calibration	152
B.1.4	Extended Model	153
B.2	Details Omitted From Main Text	155
B.2.1	Details of Isolating Each Effect	155
B.2.2	Cross-Markets Analysis With 10% REO Discount	155
B.3	Judicial vs. Non-Judicial States	157
B.4	Data Sources and Calculations	158

C	Supplement to Chapter 3	163
C.1	No Nash Equilibria with Asymmetric Behavior within Countries	163
D	Supplement to Chapter 4	166
D.1	Existence and Uniqueness of a Non-trivial Symmetric Solution of the First-Order Conditions	166
D.2	Nonexistence of Asymmetric Nash Equilibria	167

List of Tables

1.3.1 Variables in Structural Model of the Firm	8
1.6.1 Calibration of Model Parameters	31
1.6.2 Calibration of Model Parameters by Credit Rating	32
1.6.3 Credit Spreads on 10-Year Maturity Debt (bps)	34
1.6.4 Credit Spreads on 20-Year Maturity Debt (bps)	35
1.6.5 Cumulative Default Probabilities - Long Maturity	36
1.6.6 Credit Spreads on 4-Year Maturity Debt (bps)	39
1.7.1 Cumulative Default Probabilities - Short Maturity	42
1.7.2 Baa-Rated Credit Spreads with Multiscale Stochastic Volatility	44
1.7.3 Baa-Rated Cumulative Default Probabilities with Multiscale Stochastic Volatility	45
1.7.4 Baa-Rated Credit Spreads with Multiscale Stochastic Volatility	47
2.2.1 MSA Summary Statistics	56
2.2.2 Cross MSA Regressions on the Impact of the Size of the Bubble and Its Interaction With High LTV	57
2.3.1 Variables in Housing Market Model	60
2.5.1 Variables Used In Extended Model	75
2.6.1 Calibrated Parameter Values	79
2.6.2 Sensitivity Analysis: Time Out of Market For Renters	80
2.6.3 Sensitivity Analysis: Loan Balance Distribution and REO Discount	81
2.8.1 Effects of Foreclosure Policies	89
B.3.1 Judicial vs. Non-Judicial States	157

List of Figures

1.5.1 Variance Betas and Financial Distress	29
1.6.1 Debt Vegas: Endogenous and Exogenous Default	37
1.7.1 Total Debt Vegas and Average Debt Maturity	40
2.1.1 The Role of Foreclosures in the Housing Downturn	50
2.2.1 Price and Transaction Volume in Selected MSAs With High Levels of Foreclosures	54
2.2.2 Boom vs. Bust Across MSAs	55
2.3.1 Schematic Diagram of No Foreclosure Steady State of Model	59
2.3.2 Schematic Diagram of Housing Market Model With Foreclosure	62
2.4.1 Housing Market Model: Qualitative Results	69
2.4.2 Isolating the Role of Each Effect	72
2.5.1 Extended Model Schematic Diagram	74
2.7.1 National Calibration With Permanent Price Drop	83
2.7.2 Cross-MSA Simulations vs. Data	84
2.8.1 Policy: Various Sized Backlogs	90
B.2.1 Cross-MSA Simulations vs. Data: 10% REO Discount	156

For my parents.

Chapter 1

Stochastic Volatility and Asset Pricing Puzzles

1.1 Introduction

An enormous body of work in financial econometrics has documented that volatility in market returns is time-varying and has attempted to model its dynamics.¹ Chacko and Viceira (2005) and Liu (2007) point out that this time-varying volatility implies long-term investors will value assets for their ability to hedge against market volatility risk in addition to their expected returns and other hedging properties. In market equilibrium, then, assets which hedge against persistent volatility risk should require lower average returns, all else equal.² Campbell, Giglio, Polk, and Turley (2012) formally model this point, constructing an intertemporal capital asset pricing model (ICAPM) incorporating stochastic volatility and documenting the existence of low-frequency movements in market volatility.

This paper builds on these insights by constructing a real options, term-structure model of the firm that includes persistent, negatively priced shocks to volatility. In contrast to Campbell et al. (2012), the primary focus is to understand how differences in financing, productivity, and investment options across firms and their relationship with volatility generate cross-sectional variability in asset prices. The paper makes two key contributions. I first demonstrate that the model's qualitative asset pricing predictions offer new insights on

¹Much of the literature provides models similar to the ARCH/GARCH models of Engle (1982) and Bollerslev (1986) in which volatility is function of past return shocks and its own lags. More recent literature has used high-frequency data to directly estimate the stochastic volatility process. Papers include Barndorff-Nielsen and Shephard (2002), Bollerslev and Zhou (2002), and Andersen et al. (2003).

²This will be the case if, in particular, asset prices are priced according to the first-order conditions of a long-term, Epstein-Zin representative investor with a coefficient of relative risk aversion greater than one.

the value premium, financial distress, and momentum puzzles in cross-sectional equity pricing and generate new testable predictions.³ I then calibrate key parameters and analyze the model’s quantitative implications for debt pricing. The model, when calibrated to match historical leverage ratios and recovery rates, can generate empirically observed levels of credit spreads and default probabilities across ratings categories *if and only if* equityholders are allowed to optimally decide when to default. Thus, the model quantitatively delivers a resolution of the credit spread puzzle documented by Huang and Huang (2003) that existing structural models with only a single sourced price of risk significantly underpredict credit spreads, especially for investment grade debt, when matching historical recovery rates, leverage ratios, and empirical default frequencies.

To fix ideas and elucidate the key intuitions at work in the model, consider a firm with debt and suppose that the equityholders can optimally choose to either default on the firm’s obligations or to continue financing the firm by issuing new equity. Then part of the value of the firm is the equityholders’ default option. By the standard logic of option theory, an increase in volatility should raise the value of this embedded option. For example, if uncertainty about demand for the firm’s product increases, the potential downside to equityholders is capped by limited liability, while there is substantial upside potential.⁴

However, an increase in volatility has an offsetting effect. Since the firm maintains a stationary capital structure, it is exposed to rollover risk. At every point in time, the firm retires a given quantity of principal and issues new debt with principal exactly equal to this amount. Given that new debt must be issued at market value, the firm will then either face a cash shortfall or windfall as part of its flow dividend depending on the current price. Increases in volatility amplify this risk and lower the expected present value of future dividends accruing to equityholders.⁵ Therefore, an increase in volatility affects equityholders negatively as well as positively.

If the firm is currently in financial distress, i.e. close to default, then most of the equity value is comprised of the default option. At these high levels of default probability, the increased option value from higher volatility dominates the exacerbated rollover risk and financially distressed firms serve as a hedge against volatility in the market, rising in market value when volatility increases. Conversely, the default option only constitutes a small fraction of the equity value of a healthy firm, such that the effect on rollover risk dominates

³The value premium puzzle, due to Basu (1977,1983) and Fama and French (1993), refers to the greater risk-adjusted return of high book-to-market stocks over low book-to-market stocks. The momentum puzzle, attributed to Jagadeesh and Titman (1993), is the finding that a portfolio long winners and short losers generates positive CAPM alpha. The financial distress puzzle, uncovered by Dichev (1998) and Campbell, Hilscher, and Szilagyi (2008), refers to the positive CAPM alpha on a portfolio long financially healthy firms and short firms close to default.

⁴Limited liability has been studied in the corporate finance literature in papers such as Hellwig (1981), Innes (1990), Laux (2001), and Biais et al. (2012).

⁵Increases in volatility decrease the expected present value of future dividends since the value of newly issued debt, and hence the flow dividend of the firm, are concave in the firm’s productivity level.

and increases in volatility lower the equity value. According to the logic of the ICAPM, this implies that healthy firms should have higher variance risk premia than financially distressed firms. Empirically, this mechanism would manifest itself as a positive CAPM alpha in a portfolio which is long healthy firms and short financially distressed firms, hence resolving the financial distress puzzle.

In a similar fashion, the model also predicts a positive CAPM alpha of a portfolio which is long recent winners and short recent losers, consistent with the momentum anomaly. More specifically, the model predicts that abnormal risk-adjusted momentum profits should be concentrated among firms with low credit ratings, which has been empirically confirmed by Avramov, Chordia, Jostova, and Philipov (2007). Intuitively, if a firm in financial distress experiences a string of positive returns, then the variance beta of the equity falls as the firm's health improves and the importance of the equityholders' default option decreases. On the other hand, the variance beta of a firm with low credit rating experiencing a string of negative returns should increase as it is likely the firm is becoming more financially distressed. Since shocks to volatility are negatively priced, the variance risk premium and hence risk-adjusted returns of low-credit-rating winners are higher than those of low-credit-rating losers.

The model's cross-sectional predictions for firms which differ in their book-to-market ratios are more nuanced. In the model, firms can have different book-to-market ratios in a variety of ways. I focus on two.⁶ First, consider two firms and suppose their book leverage ratios and ratios of growth options to book value are the same. Then, by scale invariance, the firm with lower productivity will have the higher book-to-market ratio. If this firm is also financially distressed, then the firm with the *lower* book-to-market ratio, i.e. higher productivity, should earn greater risk-adjusted returns since its variance beta would be lower.

However, if both firms are healthy and the average maturity of their debt is short, then it is the firm with lower productivity and *higher* book-to-market ratio that earns greater risk-adjusted returns. To see this, note that among healthy firms, exposure to volatility is largely due to rollover risk, which becomes more important as a component of equity valuation as the average maturity of the firm's debt falls. But since the market value of debt is more sensitive to volatility at lower productivities, increases in volatility exacerbate rollover risk more at lower productivities. Therefore, among two healthy businesses, the firm with lower productivity and higher book-to-market ratio has greater negative exposure to volatility and thus commands higher risk premia. In this fashion, the model can generate both a value premium *and* a growth premium.

Second, firms within the model having the same book leverage ratios and productivity can still differ in their book-to-market ratios if one, the growth firm, has a higher ratio of investment options to book value. These embedded options operate in a similar manner for growth firms as does the default option for

⁶In addition to the mechanisms discussed, book-to-market ratios can also differ if firms have different levels of asset volatility or different book leverage ratios.

financially distressed firms. Since the values of the growth options increase with volatility, all else equal, a firm with a higher ratio of growth options to book value will serve as a better hedge against volatility risk in the market and should earn a lower variance risk premium. This channel therefore generates a value premium.

Turning to fixed income, note that while equity is long the option to default, debt is short. An increase in volatility raises the probability of default, which decreases the value of debt. As a result, the required return on debt should be higher in a model with stochastic volatility than a model in which volatility remains constant. This effect lowers prices and increases credit spreads. However, while this intuition is strong and important, quantitative calibration of the model illustrates that it is itself insufficient to match empirical credit spreads across ratings categories. If the default barrier is set exogenously, then the model simply creates a credit spread puzzle in the other direction. It is able to better explain the credit spreads on investment grade debt, but ends up substantially *overpredicting* the credit spreads on junk debt. This result is a reflection of the fact that existing structural models with constant volatility actually do a reasonably good job predicting the credit spreads of junk debt.

The paper demonstrates that it is exactly the interaction of stochastic volatility *and* endogenous default which allows the model to quantitatively match credit spreads across credit ratings and therefore resolve the credit spread puzzle. The reasons are twofold. First, as has been discussed, when volatility increases, it raises the value of the equityholders' default option. If equityholders optimally decide when to default, they respond to an increase in volatility by postponing default. This channel ameliorates the adverse effects of higher volatility on debt value and, in fact, can reverse the sign if the firm is sufficiently financially distressed. That is, the debt of a firm extremely close to default actually benefits from an increase in volatility and thus hedges volatility risk.⁷ In a precise sense to be made clear, the price of junk debt reflects these hedging properties such that the model generates lower credit spreads than a model with exogenous default.

Second, the simple fact that volatility is stochastic raises the value of the default option such that at all volatility levels, the endogenous default barrier is lower than in a model with constant volatility. This once again tends to raise debt values, but the effect of the shift is strongest for junk debt since it is closest to default and the future is discounted. With these two effects present, the calibrated model is able to quantitatively match well the target credit spreads and historical default probabilities across all ratings categories for intermediate and long maturity debt, holding *fixed* the market price of variance risk.

The model is less successful at shorter maturities. While still providing a substantial improvement over a model with constant volatility, it accounts for less of the empirically observed credit spreads than at longer

⁷This result implies that the value of the firm increases with volatility close to the default boundary, since the equity of financially distressed firms also increases with volatility.

maturities. The calibration further demonstrates, though, that some of this underprediction is due to the model significantly understating credit risk at these maturities. In two extensions, I consider additions to the model which allows it to generate higher credit risk at short maturities, while only marginally affecting credit risk at longer maturities. In the first, I include a fast-moving volatility time scale in addition to the slow one. In the second, I include rare disasters in the form of low-frequency jumps in the firm productivity process.

That stochastic volatility has not received much attention in structural credit modeling is likely due to the considerable technical difficulties involved, a fact pointed out by Huang and Huang (2003). To overcome the technical hurdles, I employ novel perturbation techniques from mathematical finance and physics to construct accurate, approximate asymptotic series expansions of contingent claim valuations. The fundamental assumption which makes this methodology operative in the primary model is that volatility is slowly-moving and persistent. While the econometric literature has uncovered multiple time scales in volatility dynamics, note that it is exactly these persistent fluctuations which long-run investors should care about and should therefore be significantly priced in equilibrium. The key advantages of this approach are twofold. First, it provides analytic tractability by transforming the solution of difficult partial differential equations problems with an unknown boundary to recursively solving a standard sS problem, and then a straightforward ordinary differential equations problem involving key comparative statics of the model in which volatility remains constant. Second, an application of the Feynman-Kac formula provides a useful probabilistic interpretation of the first-order correction terms which allows one to cleanly see the various mechanisms at work in the model.

The paper proceeds as follows. Section 1.2 briefly reviews the relevant literature. Section 1.3 introduces the baseline model and provides characterizations of contingent claims valuations as solutions to appropriately defined partial differential equations problems. Section 1.4 discusses the perturbation methodology used to solve the problems. Section 1.5 considers the qualitative implications of the model for the equity puzzles and provides testable predictions. Section 1.6 calibrates the model and demonstrates how it quantitatively resolves the credit spread puzzle. Section 1.7 considers extensions to the baseline model. Finally, Section 1.8 concludes.

1.2 Literature Review

In addition to the ICAPM of Campbell et al. (2012), this paper is connected to a broader literature recognizing the asset pricing implications of stochastic volatility. The long-run risks model of Bansal and Yaron (2004) incorporates time-varying consumption volatility into a consumption-based asset pricing

framework. Later calibrations of the model by Beeler and Campbell (2012) and Bansal, Kiku, and Yaron (2012) emphasize the importance of this feature in delivering empirically reasonable results. Coval and Shumway (2001), Ang, Hodrick, Xing, and Zhang (2006), Adrian and Rosenberg (2008), and Carr and Wu (2009) provide empirical evidence that innovations in market volatility are priced risk factors in the cross-section of stock returns. This paper differs from those above by explicitly constructing a structural model of the firm to understand how differences across firms can explain cross-sectional asset pricing patterns observed in the data for both equity and debt.

Previous papers have offered non-behavioral explanations for the equity puzzles considered in this paper. Gomes, Kogan, and Zhang (2003) construct a model in which book-to-market serves as a proxy for the systematic risk of assets in place. Zhang (2005) demonstrates how a model with both idiosyncratic productivity and convex capital adjustment costs can generate a value premium. Cooper (2006) is similar but includes non-convex adjustment costs. Carlson, Fisher, and Giammarino (2004) generate book-to-market effects in a model with operating leverage. Garlappi and Yan (2011) argue that the financial distress puzzle can be accounted for by a model of partial shareholder recovery.

One consistent feature in all of this work is that there is a single source of priced risk. Therefore, in explaining the value premium puzzle, for instance, the models generate higher conditional market betas of value firms than growth firms, contradicting empirical evidence. This paper differs by including a second source of priced risk, such that abnormal risk-adjusted returns can be explained by variance betas rather than market betas. In this fashion, the study is similar to Papanikolaou (2011), which includes investment shocks as a second source of priced risk to explain the value premium puzzle.

The paper is part of a large literature on structural credit modeling. Important contributions are Merton (1974), Black and Cox (1976), Leland (1994a,1994b), Longstaff and Schwartz (1995), Anderson, Sundaresan and Tychon (1996), Leland and Toft (1996), Leland (1998), Duffie and Lando (2001), and Collin-Dufresne and Goldstein (2001). Features of these models include stochastic interest rates, endogenous default, shareholder recovery, incomplete accounting information, and mean-reverting leverage ratios. Yet, as Huang and Huang (2003) show, these models cannot jointly produce historical default probabilities and realistic credit spreads. My work is most similar to Hackbarth, Miao, Morellec (2006), Chen, Collin-Dufresne (2009), and Chen (2010) in analyzing how business cycle variation in macroeconomic conditions impacts credit spreads. My work differs from these in allowing for both independent diffusive movements in volatility as well as endogenous default and, moreover, analyzing the credit spreads on junk debt in addition to the Aaa-Baa spread.

1.3 Structural Model of the Firm

In this section, I develop a continuous-time, real options model of the firm incorporating both stochastic volatility of the firm productivity process and strategic default by equityholders. The model will allow for an analysis of equity pricing as well as the full term structure of credit spreads. In later sections, I enrich this base framework by including multiple time scales in the volatility dynamics and rare disasters in the productivity process. Table 1.3.1 defines the model's key variables for convenient reference.

1.3.1 Firm Dynamics - Physical Measure

Set a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let \mathbf{W}_t be a Wiener process or standard Brownian motion in two dimensions under \mathbb{P} , which I will call the physical measure. Associated with this Brownian motion is a filtration \mathcal{F}_t satisfying the usual properties.⁸ Firms are value-maximizing and operate in a perfectly competitive environment. The productivity of a representative firm's capital is assumed to follow a stochastic process given by

$$dX_t/X_t = \mu dt + \sqrt{Y_t} dW_t^{(1)}, \quad (1.3.1)$$

where $W_t^{(1)}$ is a standard Brownian motion and μ is the expected growth rate of productivity under the physical measure \mathbb{P} .^{9,10} The variance of this process Y_t is itself stochastic and follows a mean-reverting Cox-Ingersoll-Ross (CIR) process under the physical measure given by¹¹:

$$dY_t = \kappa_Y (\theta_Y - Y_t) + \nu_Y \sqrt{Y_t} dW_t^{(2)}. \quad (1.3.2)$$

Here, κ_Y is the rate of mean reversion, θ_Y is the long-run mean of variance, and ν_Y controls the volatility of variance. The process $W_t^{(2)}$ is a Brownian motion which has correlation ρ_Y with the process $W_t^{(1)}$. In

⁸In particular, for $0 \leq s < t$, $\mathcal{F}_s \subseteq \mathcal{F}_t$. For each $t \geq 0$, the process \mathbf{W}_t is \mathcal{F}_t -measurable. Finally, for $0 \leq t < \tau$, the increment $\mathbf{W}_\tau - \mathbf{W}_t$ is independent of \mathcal{F}_t . See definition 3.3.3. of Shreve (2004).

⁹The assumption of stochastic cash flows differs from Leland (1994) and many other models of corporate debt which directly specify a stochastic process for the value of the unlevered firm. As discussed by Goldstein et al. (2001) this assumption has several advantages with regards to the modeling of tax shields and calibration of the risk neutral drift. It is also easier to incorporate growth options into such a framework.

¹⁰For technical reasons, I actually require the diffusion for cash flow to be given by:

$$dX_t/X_t = \mu dt + f(Y_t) dW_t^{(1)},$$

where $f(\cdot)$ is a smooth, positive function which is bounded and bounded away from zero. Set $f(Y_t) = \sqrt{Y_t}$ over a sufficiently large compact interval and use bump functions (also known as mollifiers) to guarantee smoothness at the boundaries.

¹¹I could have alternatively considered a mean-reverting Ornstein-Uhlenbeck process in which the volatility of variance does not scale with the current level. I choose a CIR process to be consistent with much of the empirical literature on index options.

Table 1.3.1: Variables in Structural Model of the Firm

Variable	Description
State Variables	
X_t	Asset productivity
Y_t	Asset variance
Contingent Claims	
U^a, U	Value of assets in place
E^a, E	Equity values
\tilde{d}^a, \tilde{d}	Value of newly issued debt
D^a, D	Total debt values
V^a, V	Total debt value plus equity value
u^a, u	Cumulative survival probabilities
Parameters	
r	Riskfree rate
μ, g	Expected productivity growth rates
κ_Y, κ_Z	CIR rates of mean-reversion
θ_Y, θ_Z	Long-run variances
ν_Y, ν_Z	Volatility of variance
ρ_Y, ρ_Z	Correlations between productivity and variance
ρ_{YZ}	Correlation between variance shocks
K, K^a	Capital stocks of mature, young firms
λ	Arrival rate of rare disasters
I	Cost of investment
ϕ	Corporate tax rate
ξ	Bankruptcy costs
P	Total principal
C	Total coupon
m	Rollover rate of debt
$\pi_X(Y_t)$	Asset risk premium
$\Gamma(Y_t)$	Market price of variance risk

Note: This table defines the key variables, parameters, and notation used throughout the paper. A subscript a refers to young firms. The subscript Y refers to the slow-moving volatility factor and the subscript Z refers to the fast-moving volatility factor. The parametric specification for the market price of volatility risk is given by $\Gamma(Y_t) = \Gamma_0 \sqrt{Y_t}$.

particular, I have that

$$\begin{pmatrix} W_t^{(1)} \\ W_t^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho_Y & \sqrt{1 - \rho_Y^2} \end{pmatrix} \mathbf{W}_t, \quad (1.3.3)$$

where \mathbf{W}_t is the Wiener process defined above.

The firm has initial capital $K^a < 1$ and can irreversibly expand its productive capacity to $K = 1$ at the discretion of the equityholders. Exercising the growth option costs $I > 0$.¹² I define firms with capital K^a to be “young” or “growth” firms and those with capital K as “mature” firms. Finally, firms pay taxes on their income at the corporate rate ϕ so that the flow of after-tax profits of the unlevered firm at time t is

¹²The choice of $K = 1$ is simply a normalization reflecting the scale invariance of the model. Investment occurs discretely, which allows me to solve for the contingent claims of mature and young firms recursively.

given by

$$(1 - \phi) X_t K_t. \quad (1.3.4)$$

As is standard, this profit function reflects optimal choices by the firm in all other variable inputs such as labor and raw materials.

1.3.2 Firm Dynamics - Risk Neutral Measure

I assume the existence of an equivalent martingale measure \mathbb{P}^* under which contingent claims will be priced.

In particular, I suppose there exist processes $\gamma_1(Y_s), \gamma_2(Y_s) \in \mathcal{L}^2$ such that the process

$$\mathbf{W}_t^* = \mathbf{W}_t + \int_0^t \begin{pmatrix} \gamma_1(Y_s) \\ \gamma_2(Y_s) \end{pmatrix} ds \quad (1.3.5)$$

is a martingale under \mathbb{P}^* . I define the *market price of risk* as $\gamma_1(Y_t) > 0$ and the *market price of variance risk* as

$$\Gamma(Y_t) = \rho_Y \gamma_1(Y_t) + \sqrt{1 - \rho_Y^2} \gamma_2(Y_t). \quad (1.3.6)$$

These definitions are motivated by the fact that due to Girsanov's theorem, under the risk-neutral measure, the dynamics of productivity and variance are given by

$$dX_t = \left(\mu - \gamma_1(Y_t) \sqrt{Y_t} \right) X_t dt + \sqrt{Y_t} X_t dW_t^{(1)*} \quad (1.3.7)$$

$$dY_t = \left(\kappa_Y (\theta_Y - Y_t) - \Gamma(Y_t) \nu_Y \sqrt{Y_t} \right) dt + \nu_Y \sqrt{Y_t} dW_t^{(2)*}, \quad (1.3.8)$$

where $W_t^{(i)*}$ is defined as in (1.3.3) replacing \mathbf{W} with \mathbf{W}^* .

Since I am considering the setting of random, non-tradeable volatility, I am not in a complete markets framework and therefore the equivalent martingale measure is not unique. In fact, there exists a family of pricing measures parameterized by the market price of variance risk. Note that the market price of variance risk will not be zero if there is correlation between movements in variance and movements in productivity. That is, the first term in (1.3.6) will be nonzero and have the same sign as ρ_Y . However, theory suggests that independent fluctuations in variance may also be priced. As discussed in the introduction, if long-term investors see increases in volatility as deteriorations in the investment opportunity set, then the price $\gamma_2(Y_t)$ will be negative.

I define the *asset risk premium* to be $\pi_X(Y_t) = \gamma_1(Y_t) \sqrt{Y_t}$ and assume it to be independent of current variance Y_t . This allows me to define $g = \mu - \pi_X$ as the expected growth rate of productivity under the

risk-neutral measure. Finally, I will refer to the expression $\nu_Y \sqrt{Y_t} \Gamma(Y_t)$ as the *variance risk premium*.¹³ As this object will be of central importance in the paper, it therefore warrants further discussion. The variance risk premium is the expected excess return over the riskfree rate on any asset with zero market beta and a variance beta equal to one, where these betas are respectively defined as¹⁴

$$\beta_X = x \frac{\partial \log(\cdot)}{\partial x} \quad (1.3.9)$$

$$\beta_Y = \frac{\partial \log(\cdot)}{\partial y}. \quad (1.3.10)$$

For instance, it gives the expected excess return on a portfolio of delta-neutral straddles or a portfolio of variance swaps constructed to have a unit variance beta. Coval and Shumway (2001) estimate the expected excess returns on delta-neutral straddles and find them to be negative. Carr and Wu (2009) document a similar fact for variance swaps. These empirical findings are supportive of a negative variance risk premium as predicted by the ICAPM.

For my calibration and quantitative analysis, I will adopt the following parametric specification for the market price of variance risk:

$$\Gamma(Y_t) = \Gamma_0 \sqrt{Y_t}, \quad (1.3.11)$$

which leads to a variance risk premium that is affine in the current variance Y_t . Beginning with Heston (1993), this has been the standard assumption in both the option pricing and financial econometrics literature. Its appealing feature is it implies the variance dynamics follow a Cox-Ingersoll-Ross (CIR) process under both the risk-neutral and physical measures. The only difference is the rate of mean-reversion. Note, however, that the theory which follows will not rely at all on this parameterization and holds for any arbitrary specification $\Gamma(Y_t)$.

The value of the assets in place is given by the discounted present value under the risk-neutral measure of future after-tax profits generated by installed capital. Denoting this value by U_t for mature firms and letting expectations under \mathbb{P}^* conditional on \mathcal{F}_t be denoted by $\mathbb{E}_t^*(\cdot)$, I have

$$U_t(x) = \mathbb{E}_t^* \left[\int_t^\infty e^{-r(s-t)} (1 - \phi) X_s ds \mid X_t = x \right] = \frac{(1 - \phi)x}{r - g}, \quad (1.3.12)$$

¹³The distinction between prices of risk and risk premia are confused in the literature. I take care here to be explicit about their definitions.

¹⁴These definitions reflect the use of Ito's lemma in calculating the required excess return of an asset under the physical measure.

which is simply the Gordon growth formula given that after-tax profits grow at the rate g under the risk-neutral measure. From this I can see that

$$dU_t/U_t = gdt + \sqrt{Y_t}dW_t^{(1)*}, \quad (1.3.13)$$

so that the value of assets in place and productivity share the same dynamics under \mathbb{P}^* . Finally, one can show that

$$\frac{dU_t + (1 - \phi)X_t}{U_t} = rdt + \sqrt{Y_t}dW_t^{(1)*}, \quad (1.3.14)$$

which implies that the expected risk-neutral total return on the assets in place, including the dividend payment, is equal to the riskless rate as required by the existence of an equivalent martingale measure. The value of assets in place for young firms, denoted by U_t^a , is calculated similarly and is provided in Appendix A.2.

1.3.3 Capital Structure

Firms are financed with both debt and equity issues. I employ the exponential model of Leland (1994a) and Leland (1998) to describe the capital structure of the firm. This modeling device will allow me to analyze the full term structure of credit spreads, while at the same time maintain tractability.¹⁵ The firm adopts a stationary debt structure with total principal P and total coupon C . The firm continuously rolls over debt at the fractional rate m . That is, at every point in time, debt with principal equal to mP matures and is replaced with new debt of equal coupon, principal, and seniority to maintain stationarity. Let p denote the principal on newly issued debt and c the coupon. Moreover, let $p(s, t)$ and $c(s, t)$ be the principal and coupon outstanding at time t for debt issued at date $s \leq t$. Since the firm retires the principal of all vintages at fractional rate m , the principal and coupon of each vintage declines exponentially with time:

$$p(s, t) = e^{-m(t-s)}p \quad (1.3.15)$$

$$c(s, t) = e^{-m(t-s)}c. \quad (1.3.16)$$

¹⁵ An alternative term structure model is given by Leland and Toft (1996). However, such a model would require me to solve partial differential equations for the debt and equity values. A significant advantage of the exponential model is that, as will be shown, approximate equity and debt values can be derived by solving only ordinary differential equations.

Integrating over all vintages at time t gives the total principal and coupon of the firm:

$$P = \int_{-\infty}^t p(s, t) ds = p \int_{-\infty}^t e^{-m(t-s)} ds = p/m \quad (1.3.17)$$

$$C = \int_{-\infty}^t c(s, t) ds = c \int_{-\infty}^t e^{-m(t-s)} ds = c/m. \quad (1.3.18)$$

Thus, the principal and coupon of newly issued debt is always equal to a fraction m of the total principal and coupon in the capital structure of the firm.

Now, note that for a time s vintage, the fraction of currently outstanding debt principal retired at time $t > s$ is given by $me^{-m(t-s)}$. This implies that the average maturity of debt M is given by

$$M = \int_s^{\infty} t \left(me^{-m(t-s)} \right) dt = 1/m. \quad (1.3.19)$$

In other words, the inverse of the fractional rollover rate is a measure of the average maturity of the firm's debt. Note that the limiting case $m = 0$ corresponds to the Leland (1994b) model of consol debt.

Of course, the price of newly issued debt will reflect current state variables in the market. This exposes the equityholders of the firm to *rollover risk*. The firm will face either a cash windfall or shortfall depending on whether debt is currently priced above or below par. Additional equity must be issued if there is a cash shortfall. Since interest payments provide a tax shield, the total flow payment to equityholders of mature firms is therefore given by

$$(1 - \phi) [X_t - C] + \tilde{d}(X_t, Y_t) - p, \quad (1.3.20)$$

where $\tilde{d}(X_t, Y_t)$ is the value of newly issued debt. The first term reflects the flow operating profits generated by the the firm as well as the corporate tax rate and the tax shield, while the final two terms reflect the rollover risk faced by the equityholders.

Equityholders optimally decide when to default on their debt obligations and may issue additional equity to finance coupon payments if current cash flow is insufficient to meet their obligations. In the event of default, equityholders receive nothing and the value of their claims is zero. Debtholders receive the assets of the firm according to their vintage; however, a fraction ζ of the total value of assets in place is lost due to bankruptcy costs.¹⁶

Finally, I will assume for simplicity that investment spending is financed entirely with equity. I now turn to the valuation of the firm's contingent claims.

¹⁶Bankruptcy costs include the direct costs of the legal proceedings, but also indirect costs such as losses of specialized knowledge and experience, reductions in trade credit, and customer dissatisfaction. In practice, the indirect costs of bankruptcy may be of an order of magnitude larger than the direct costs.

1.3.4 Equity Valuation

Contingent claims are priced according to the risk-neutral measure. Given equation (1.3.20), the value of equity for a mature firm is given by

$$E(x, y) = \sup_{\tau \in \mathcal{T}} \mathbb{E}_t^* \left[\int_t^\tau e^{-r(s-t)} \left\{ (1 - \phi)(X_s - C) + \tilde{d}(X_s, Y_s) - p \right\} ds \right], \quad (1.3.21)$$

where \mathcal{T} is the set of $\{\mathcal{F}_t\}$ -stopping times. I denote the optimal stopping time, i.e. time of default, by τ^B . Intuitively, the equity value of a mature firm is simply the risk-neutral expected discounted present value of all future dividends accruing to equityholders given that the point of default is chosen optimally.

Note that simple Monte-Carlo computation of this value is not feasible due to the unknown optimal stopping time. Additionally, one cannot implement the recursive least-squares Monte-Carlo (LSM) procedure of Longstaff and Schwartz (2001) since there is no terminal date.¹⁷ Instead, I will seek a partial differential equations characterization of the equity valuation. Specifically, I show that the equity value of the mature firm (1.3.21) can be given as the solution to a Dirichlet/Poisson free boundary problem.

Theorem 1.3.1 *The equity value of a mature firm $E(x, y)$ is the solution to*

$$(1 - \phi)(x - C) + \tilde{d}(x, y) - p + \mathcal{L}_{X,Y}E = rE \quad \text{for } x > x_B(y) \quad (1.3.22a)$$

$$(1 - \phi)(x - C) + \tilde{d}(x, y) - p + \mathcal{L}_{X,Y}E \leq rE \quad \text{for } x, y > 0 \quad (1.3.22b)$$

$$E(x, y) = 0 \quad \text{for } x < x_B(y) \quad (1.3.22c)$$

$$E(x, y) \geq 0 \quad \text{for } x, y > 0 \quad (1.3.22d)$$

$$E(x_B(y), y) = 0 \quad (1.3.22e)$$

$$\lim_{x \rightarrow \infty} E(x, y) = U(x) - \frac{C + mP}{r + m} + \frac{\phi C}{r} \quad (1.3.22f)$$

$$\frac{\partial E}{\partial x} \Big|_{x=x_B(y)} = 0 \quad (1.3.22g)$$

$$\frac{\partial E}{\partial y} \Big|_{x=x_B(y)} = 0, \quad (1.3.22h)$$

where $\mathcal{L}_{X,Y}$ is the linear differential operator given by:

$$\mathcal{L}_{X,Y} = gx \frac{\partial}{\partial x} + \frac{1}{2}yx^2 \frac{\partial^2}{\partial x^2} + (\kappa_Y(\theta_Y - y) - \Gamma(y)\nu_Y\sqrt{y}) \frac{\partial}{\partial y} + \frac{1}{2}\nu_Y^2 y \frac{\partial^2}{\partial y^2} + \rho_Y \nu_Y y x \frac{\partial^2}{\partial x \partial y}, \quad (1.3.23)$$

and $x_B(y)$ is a free boundary to be determined.

¹⁷Essentially, the methodology uses recursive Monte-Carlo simulation backwards in time along with least-squares regression to compute conditional expectations and compares the expected continuation value against the intrinsic value at each discrete time step to approximate the exercise boundary.

Proof. See Appendix A.1.1. ■

Here $x_B(y)$ is the point at which equityholders optimally default on their debt. That is, the optimal stopping time $\tau^B = \min\{t : (X_t, Y_t) = (x_B(Y_t), Y_t)\}$, which is clearly adapted to the filtration \mathcal{F}_t and is thus well-defined. Crucially, note that the boundary at which default occurs can depend on the current volatility.

The other conditions in the above system are intuitive. By the multidimensional Ito's formula, I can write equation (1.3.22a) as

$$\frac{(1 - \phi)(x - C) + \tilde{d}(x, y) - p}{E} + \frac{E^*[dE]}{E} = r, \quad (1.3.24)$$

which simply says that the expected return on equity, given by the dividend yield plus the expected capital gain, must be equal to the risk-free rate under the risk-neutral measure if the firm is not in default. Equation (1.3.22b) implies that in the stop region the expected return from continuing operations must be less than or equal to the riskless rate. Otherwise, stopping would not be optimal. Equation (1.3.22e) is the usual value-matching equation which states that at the point of default the value of equity must be equal to zero. If the value, for instance, were positive then the continuation value of equity would be larger than the value of equity under default, and thus default should be optimally postponed. As $X_t \rightarrow \infty$, the probability of default in finite time approaches zero, and therefore the equity value is simply given by the present value of the assets in place U_t , minus the present value of future debt obligations $(C + mP)/(r + m)$, plus the present value of the tax shield $\phi C/r$. This logic yields the limiting condition (1.3.22f).

Finally, equations (1.3.22g) and (1.3.22h) provide the smooth-pasting conditions for the problem. These are standard for optimal stopping problems in which the stochastic process follows a regular diffusion. To see why this should be the case, note that if the equity value were not smooth across the free boundary in both variables, there would be a kink in the valuation at the point of default. The nature of diffusion is such that this kink would imply defaulting slightly earlier or later would be optimal. For instance, suppose that the kink were convex. Loosely speaking, within a short period of time the diffusion would be equally likely to be on either side of the kink and since the slope into the continuation region is positive, there is positive expected value to waiting. The payoff from stopping immediately is zero and thus it would be optimal to postpone default.

Similarly, letting $E^a(x, y)$ denote the equity value of young firms, it solves the optimal stopping problem:

$$E^a(x, y) = \sup_{\tau', \tau'' \in \mathcal{T}} \mathbb{E}_t^* \left[\begin{aligned} & \left\{ \int_t^{\tau' \wedge \tau''} e^{-r(s-t)} (1 - \phi)(K_a X_s - C) \tilde{d}(x, y) - p \right\} ds \\ & + 1[\tau' < \tau''] (E(X_{\tau'}, Y_{\tau'}) - I) \end{aligned} \right], \quad (1.3.25)$$

where $\tau' \wedge \tau'' = \min(\tau', \tau'')$. Here τ' is a stopping time which indicates exercising the growth option and τ'' is a stopping time which indicates default. If the firm exercises its growth option prior to default, then the value of the equity becomes $E(X_{\tau'}, Y_{\tau'}) - I$, which gives the second term in the expression above. The equity value of a young firm can be described as the solution to an appropriate free boundary problem just as with mature firms. Basically, the limiting condition (1.3.22f) is removed and there are value-matching and smooth-pasting conditions at the boundary for exercise of the growth option. It is characterized explicitly in Appendix A.2.

1.3.5 Debt Valuation

As is evident from the equations above, solving the equityholders' problem requires the value of newly issued debt. This is due to the assumptions on the capital structure of the firm and the rollover risk inherent in the flow dividend to the equityholders. Let $d(t)$ be the value at date t of debt issued at time 0 for a mature firm.¹⁸ Then according to risk-neutral pricing, this valuation is given by

$$d(t) = \mathbb{E}_t^* \left[\begin{aligned} & \int_t^{\tau^B} \{e^{-r(s-t)} e^{-ms} (c + mp)\} ds \\ & + e^{-r(\tau^B-t)} \left(e^{-m\tau^B} p/P \right) (1 - \xi) U_{\tau^B} \end{aligned} \right]. \quad (1.3.26)$$

Note that the payments to the debtholders are declining exponentially and are therefore time dependent. Furthermore, the debtholders' claim on the assets of the firm will also depend on the point in time at which bankruptcy occurs. In general, this time-dependency would indicate that a partial differential equation involving a time derivative would need to be solved to find $d(t)$. Instead multiply both sides of the equation above by e^{mt} . Noting that $p/P = m$ yields

$$e^{mt} d(t) = \mathbb{E}_t^* \left[\begin{aligned} & \int_t^{\tau^B} \{e^{-(r+m)(s-t)} (c + mp)\} ds \\ & + e^{-(r+m)(\tau^B-t)} m (1 - \xi) U_{\tau^B} \end{aligned} \right]. \quad (1.3.27)$$

Applying Feynman-Kac, I then have the following result:

Theorem 1.3.2 *The value of the date 0 debt vintage at time t is given by $d(t) = e^{-mt} \tilde{d}(X_t, Y_t)$ where*

¹⁸Note that we could have considered any vintage of debt. Analyzing the date 0 vintage is without loss of generality.

$\tilde{d}(X_t, Y_t)$ is the value of the newly issued debt and satisfies

$$c + mp + \mathcal{L}_{X,Y} \tilde{d} = (r + m) \tilde{d} \quad \text{for } x > x_B(y) \quad (1.3.28a)$$

$$\tilde{d}(x_B(y), y) = m(1 - \xi) U(x_B(y)) \quad (1.3.28b)$$

$$\lim_{x \rightarrow \infty} \tilde{d}(x, y) = \frac{c + mp}{r + m}. \quad (1.3.28c)$$

The value of newly issued debt is therefore the solution to a partial differential equation that does not involve a time derivative. Rather, the exponential decline in the flow payments to the debtholders is reflected in a higher implied interest rate $r + m$. This is the key advantage of employing the Leland (1994a) model of capital structure. To find the total value of debt at any point in time, simply integrate over the value of all outstanding debt vintages:

$$D(t) = \int_{-\infty}^t e^{-m(t-s)} \tilde{d}(X_t, Y_t) ds = \frac{\tilde{d}(X_t, Y_t)}{m}. \quad (1.3.29)$$

Thus, the total value of the firm's aggregate debt is a function only of the current productivity and volatility and does not depend on time, which is consistent with the stationarity of the overall debt structure.

The debt of young firms is characterized similarly and is discussed in Appendix A.2.

1.4 Methodology

The analysis up until this point has provided time-stationary partial differential equation (PDE) characterizations of the debt and equity values. In particular, the equity value is the solution to a two-dimensional free boundary problem. It is the higher dimensional counterpart to the usual sS problem well known in economics and similar to the obstacle problem and the Stefan problem in physics.¹⁹ While it is often possible to obtain closed form solutions to sS problems with only one state variable, this is usually not feasible when the free boundary is of a higher dimension. I could attempt to directly solve the problem numerically using an appropriate discretization scheme and projected successive over-relaxation (PSOR) methods. Such techniques have been successfully applied to the solution of American option problems and involve a variant of the Gauss-Seidel method for the solution of systems of linear equations in which the relevant inequality constraint is enforced at each iteration. However, these numerical procedures can be computationally burdensome and, more importantly, would not allow me to analytically grasp the intuition and economics

¹⁹The classic Stefan problem studies phase changes in homogenous mediums and the resulting temperature distributions. The obstacle problem considers elastic membranes and solves for their equilibrium position given a fixed boundary and the constraint that they lie above a certain geometric obstacle.

underlying the results.

Instead, I utilize a semi-analytic methodology based on asymptotic expansions which allows me to generate accurate approximations to the values of the contingent claims. The method relies on making assumptions on the rate of mean-reversion of the stochastic volatility process and then using either regular or singular perturbations around the appropriate parameter in the partial differential equations to recursively solve for the formal power series expansions of the debt and equity values, as well as the default boundary. I will confine myself to first-order expansions, although higher-order terms can be calculated in a straightforward manner.

The basic principles of this method have been developed recently in the mathematical finance literature for the pricing of options under stochastic volatility.²⁰ Lee (2001) develops an approach for the pricing of European options under slow-variation asymptotics of the stochastic volatility process.²¹ In a sequence of papers, authors Fouque, Papanicolaou, Sircar, Sølna have considered fast-variation asymptotics in a variety of option pricing settings (European, barrier, Asian, etc.) and have developed perturbation procedures for stochastic volatility processes with both slow and fast variation components.²² My baseline setting will use an adaptation of the method proposed by Lee (2001) for small variation asymptotics to a free boundary problem. I will be able further extend this approach in a novel manner to allow for rare disasters in the firm productivity process. Finally, I will adapt the multiscale methods of Fouque, Papanicolaou, Sircar, Sølna, which are significantly more complicated, to consider a setting with both slowly-varying and fast-varying components of volatility.

To the best of my knowledge, the use of this methodology is novel in the economics literature outside of mathematical finance, and as I will argue below, is particularly well-suited for use by economists in a variety of economic settings. I now turn to the discussion of the existence of volatility time scales, which is important for justifying the asymptotic expansions considered, and how they are to be modeled.

1.4.1 Time Scales in Volatility Modeling

A substantial number of empirical studies have shown that market volatility appears to evolve on multiple time scales and to exhibit forms of long-run dependencies, often termed long memory. This has led to a

²⁰More broadly, these methods are based on the use of perturbation theory to solve differential equations, which has a rich history in both mathematics and physics. Perturbation methods, for instance, are widely used in the modern study of quantum mechanical models.

²¹Precisely, Lee (2001) considers the cases of both slow-variation and small-variation stochastic volatility.

²²Contributions by these authors include Fouque et al. (2003a, 2003c, 2004a, 2004b, 2006, 2011). In particular, Fouque et al. (2006) shows how defaultable bond prices can be calculated using asymptotic expansions in a Black-Cox first passage model with multiscale stochastic volatility. This paper differs by discussing the economic intuitions, including endogenous default and a stationary capital structure with rollover risk, analyzing the implications of stochastic volatility for both equity and debt, and examining the credit spread puzzle.

number of important econometric developments that attempt to move beyond the standard one-component ARCH/GARCH/EGARCH specifications. Andersen and Bollerslev (1997) and Baillie et al. (1996) introduce the Fractionally Integrated ARCH (FIGARCH) model in which autocorrelations have hyperbolic decline rather than geometric decline. Beginning with Engle and Lee (1999), a body of papers has argued that two-dimensional volatility models with both a short-run component and a long-run component show significantly better performance in matching the data.²³ Building on this work further, Calvet and Fisher (2001, 2004) have developed the Markov-switching multifractal (MSM) stochastic volatility model, allowing for an arbitrary number of time scales on which volatility can evolve. Here, volatility is the product of a large but finite number of factors, each of which are first-order Markov and have identical marginal distributions but differ in their switching probabilities. The authors find that a specification with ten time scales fits the volatility of exchange rates well, with the highest frequency component on the order of a day and the lowest component on the order of 10 years.

I adopt this perspective in my modeling of asset variance. In my primary specification, though, I will only model low frequency movements. One reason for this is that high frequency fluctuations are likely to be less important in the actual dynamics of asset volatility than market volatility, since changes in asset volatility should be driven largely by fundamentals rather than market sentiment and other transient factors. More importantly, though, it is exactly those persistent, low frequency movements in volatility which long-run investors should care about and which should be priced in equilibrium.

Rather than directly calibrating the parameter κ_Y to be small, I instead slightly modify the process for variance under the physical measure as

$$dY_t = \delta \kappa_Y (\theta_Y - Y_t) dt + \nu_Y \sqrt{\delta Y_t} dW_t^{(2)}, \quad (1.4.1)$$

where $\delta > 0$ is a small parameter. Note that this is again a CIR process exactly the same as in equation (1.3.2) except that the drift term is now multiplied by δ and the diffusion term is multiplied by $\sqrt{\delta}$. This parameter directly controls the rate of mean-reversion of the process for volatility. Since I am assuming it is small, I will say that variance is *slowly mean-reverting*. Specifically, δ scales the spectral gap of the process for Y_t , or the distance between the zero eigenvalue and the first negative eigenvalue. Using eigenfunction expansions, it is possible to show that it is exactly this spectral gap which determines the rate of mean-reversion for the process. However, denoting the invariant or long-run distribution of Y_t by Δ_Y , one can

²³Other papers include Engle and Rosenberg (2000), Alizadeh, Brandt, and Diebold (2002), Bollerslev and Zhou (2002), Fouque et al. (2003), Chernov et al. (2003), and Adrian and Rosenberg (2008).

show that

$$\Delta_Y \sim \text{Gamma} \left(\frac{2\kappa_Y \theta_Y}{\nu_Y^2}, \frac{\nu_Y^2}{2\kappa_Y} \right),$$

which is in fact independent of δ . This indicates that in the long run, the level of variability in the volatility of asset productivity does not depend on the parameter δ , even though its square root multiplies the diffusion term in (1.4.1). It is in this sense that movements in volatility are indeed *slow*, but not necessarily *small*.²⁴ Appendix A.3 provides further technical details on this modeling of volatility time scales.

In an extension of the model, I will construct a model in the spirit of Calvet and Fisher (2001, 2004) in which volatility is the product of both a low frequency component and a high frequency component.

1.4.2 Asymptotic Approximation

Given this model of volatility dynamics, I can now write the partial differential equations in (1.3.22a) and (1.3.28a) as:

$$(1 - \phi)(x - C) + \tilde{d}_\delta(x, y) - p + \left(\mathcal{L}_r^y + \sqrt{\delta} \mathcal{M}_1^y + \delta \mathcal{M}_2^y \right) E_\delta = 0 \quad (1.4.2)$$

$$c + m + \left(\mathcal{L}_{r+m}^y + \sqrt{\delta} \mathcal{M}_1^y + \delta \mathcal{M}_2^y \right) \tilde{d}_\delta = 0, \quad (1.4.3)$$

where E_δ and \tilde{d}_δ are the values of equity and newly issued debt respectively given the choice of δ and the operators $\mathcal{L}_r^y, \mathcal{M}_1^y, \mathcal{M}_2^y$ are given by

$$\mathcal{L}_r^y = gx \frac{\partial}{\partial x} + \frac{1}{2} y x^2 \frac{\partial^2}{\partial x^2} - r(\cdot) \quad (1.4.4)$$

$$\mathcal{M}_1^y = \rho_Y \nu_Y y x \frac{\partial^2}{\partial x \partial y} - \Gamma(y) \nu_Y \sqrt{y} \frac{\partial}{\partial y} \quad (1.4.5)$$

$$\mathcal{M}_2^y = \kappa_Y (\theta_Y - y) \frac{\partial}{\partial y} + \frac{1}{2} \nu_Y^2 y \frac{\partial^2}{\partial y^2}. \quad (1.4.6)$$

Here, \mathcal{L}_r^y is the time-invariant Black-Scholes operator with volatility y and riskfree rate r , \mathcal{M}_2^y is the infinitesimal generator of the CIR process, and \mathcal{M}_1^y is an operator which accounts for correlation between the processes for asset productivity and asset volatility, as well as the Girsanov transformation between the physical and risk-neutral measures.²⁵

²⁴To model the implications of small variation in asset volatility, I would specify:

$$dY_t = \delta \kappa (\theta - Y_t) dt + \delta \nu \sqrt{Y_t} dW_t^{(2)}.$$

Lee (2001) considers the implications of such a stochastic volatility model for the pricing of European options.

²⁵Note that the dynamics of volatility under the risk-neutral measure are given by:

$$dY_t = \left(\delta \kappa (\theta - Y_t) - \Gamma(Y_t) \nu \sqrt{\delta Y_t} \right) dt + \nu \sqrt{\delta Y_t} dW_t^{(2)*}$$

To solve for the contingent claims of mature firms by regular perturbation, I expand the equity value, the value of newly issued debt, and the free default boundary in powers of $\sqrt{\delta}$:

$$E_\delta(x, y) = E_0^y(x) + \sqrt{\delta}E_1^y(x) + \delta E_2^y(x) + \dots \quad (1.4.7)$$

$$\tilde{d}_\delta(x, y) = \tilde{d}_0^y(x) + \sqrt{\delta}\tilde{d}_1^y(x) + \delta\tilde{d}_2^y(x) + \dots \quad (1.4.8)$$

$$x_B(y) = x_{B,0}^y + \sqrt{\delta}x_{B,1}^y + \delta x_{B,2}^y + \dots \quad (1.4.9)$$

I then plug these asymptotic expansions into equations (1.4.2) and (1.4.3), as well as equations (1.3.22e)-(1.3.22h) and (1.3.28b)-(1.3.28c). Taylor expansions centered around $x_{B,0}(y)$ are used to appropriately expand value-matching and smooth-pasting conditions. The system is solved by collecting terms in the powers of δ and using the method of undetermined coefficients, where here I understand the coefficients to be functions.

Principal Order Terms

I begin by computing the principal order terms in the asymptotic expansions. Collecting the order one terms for newly issued debt yields the problem:

$$c + mp + \mathcal{L}_{r+m}^y \tilde{d}_0^y = 0 \quad \text{for } x > x_{B,0}^y \quad (1.4.10a)$$

$$\tilde{d}_0^y(x_{B,0}^y) = m(1 - \xi)U(x_{B,0}^y) \quad (1.4.10b)$$

$$\lim_{x \rightarrow \infty} \tilde{d}_0^y(x) = \frac{c + mp}{r + m}. \quad (1.4.10c)$$

Under closer inspection, I see that this principal order term is simply the value of newly issued debt under constant return variance y in the firm productivity process, given the fixed boundary $x_{B,0}^y$. As such, it can be solved using standard single variable ODE techniques. I obtain

$$\tilde{d}_0^y(x) = \frac{c + mp}{r + m} + \left[m(1 - \xi)U(x_{B,0}^y) - \frac{c + mp}{r + m} \right] \left\{ \frac{x}{x_{B,0}^y} \right\}^{\gamma_1}, \quad (1.4.11)$$

where γ_1 is the *negative* root of the following quadratic equation:

$$g\gamma_1 + \frac{1}{2}y\gamma_1(\gamma_1 - 1) - (r + m) = 0. \quad (1.4.12)$$

for a given choice of parameter δ .

This expression is consistent with that derived in Leland (1994a). The first term in the valuation reflects the present value of future coupon and principal payments assuming no default. The latter term takes into account the risk of default. The term $\left(x/x_{B,0}^y\right)^{-\gamma_1}$ is akin to a probability of default although this is not exactly correct. If default occurs, debtholders receive a fraction of the value of the assets of the firm but lose any future coupon and principal payments remaining. More precisely, the second component of the valuation is a perpetual digital option which pays off the term in brackets the first time the process X_t crosses the boundary $x_{B,0}$.

Intuitively, the principal order term reflects the value of newly issued debt in the limiting case $\delta = 0$. Of course, this is simply a model in which volatility is fixed at its current value. This intuition then carries over to computing the principal order terms for both the equity value and the default boundary. That is, I simply need to compute the value of equity and the default boundary in the case where volatility is fixed in time at its current level \sqrt{y} . These are given by

$$\begin{aligned} E_0^y(x) = & U(x) - \frac{C + mP}{r + m} + \frac{\phi C}{r} \\ & + \left[\frac{C + mP}{r + m} - (1 - \xi) U(x_{B,0}^y) \right] \left\{ \frac{x}{x_{B,0}^y} \right\}^{\gamma_1} \\ & - \left[\frac{\phi C}{r} + \xi U(x_{B,0}^y) \right] \left\{ \frac{x}{x_{B,0}^y} \right\}^{\gamma_2} \end{aligned} \quad (1.4.13)$$

$$x_{B,0}^y = - \frac{(C + mP) \gamma_1 / (r + m) - \phi C \gamma_2 / r}{1 - (1 - \xi) \gamma_1 - \xi \gamma_2} \frac{r - g}{1 - \phi}, \quad (1.4.14)$$

where γ_2 is the negative root to the quadratic equation:

$$g\gamma_2 + \frac{1}{2}y\gamma_2(\gamma_2 - 1) - r = 0. \quad (1.4.15)$$

These expressions are once again consistent with Leland (1994a) and are derived in Appendix A.1.2. The equity value incorporates the value of the assets in place, the present value of future debt payments, the present value tax shield, and two perpetual digital options which account for the equityholders' default option.

First-Order Correction Terms

Continuing with the recursive approach, collecting terms of order $\sqrt{\delta}$ now yields the problems to solve for the first-order correction terms. For the value of newly issued debt, this is given by

$$\begin{aligned}\mathcal{L}_{r+m}^y \sqrt{\delta} \tilde{d}_1^y &= -\sqrt{\delta} \mathcal{M}_1^y \tilde{d}_0^y \\ &= \left(A^\delta y \frac{\partial \tilde{d}_0^y}{\partial y} - B^\delta y x \frac{\partial^2 \tilde{d}_0^y}{\partial x \partial y} \right) \quad \text{for } x > x_{B,0}^y\end{aligned}\tag{1.4.16a}$$

$$\sqrt{\delta} \tilde{d}_1^y(x_{B,0}^y) = \sqrt{\delta} x_{B,1}^y \left[\frac{m(1-\xi)(1-\phi)}{r-g} - \frac{\partial \tilde{d}_0^y}{\partial x}(x_{D,0}^y) \right]\tag{1.4.16b}$$

$$\lim_{x \rightarrow \infty} \sqrt{\delta} \tilde{d}_1^y(x) = 0,\tag{1.4.16c}$$

where the constants A^δ, B^δ are defined by

$$A^\delta = \sqrt{\delta} \nu_Y \Gamma_0\tag{1.4.17}$$

$$B^\delta = \sqrt{\delta} \nu_Y \rho_Y.\tag{1.4.18}$$

Note that this is a one-dimensional inhomogeneous second-order boundary value problem in which y operates as a parameter. The source term is a function of comparative statics or Greeks of the debt principal order term. In particular, it is a function of the vega and the vanna of the debt value when volatility is held constant at the level \sqrt{y} .²⁶ Note also that the default boundary correction term $x_{B,1}^y$ appears in the boundary condition of the problem above, given by equation (1.4.16b). The problem for the first-order equity correction term takes a similar form and is given by

$$\begin{aligned}\mathcal{L}_r^y \sqrt{\delta} E_1^y &= -\sqrt{\delta} \left(\mathcal{M}_1^y E_0^y + \tilde{d}_1^y \right) \\ &= \left(A^\delta y \frac{\partial E_0^y}{\partial y} - B^\delta y x \frac{\partial^2 E_0^y}{\partial x \partial y} - \sqrt{\delta} \tilde{d}_1^y \right) \quad \text{for } x > x_{B,0}^y\end{aligned}\tag{1.4.19a}$$

$$\sqrt{\delta} E_1^y(x_{B,0}^y) = 0\tag{1.4.19b}$$

$$\lim_{x \rightarrow \infty} \sqrt{\delta} E_1^y(x) = 0.\tag{1.4.19c}$$

Once again, this is a one-dimensional second-order boundary value problem with a source term. Different from the problem for the debt value correction above, the Black-Scholes operator here has for its riskfree rate r instead of $r+m$ as in equation (1.4.19a). Furthermore, the source term for the problem involves the debt correction term, due to the rollover risk in the original problem, in addition to the vega and vanna of

²⁶Here, I define vega to be the comparative static with respect to variance y .

the equity value in the constant volatility case.

Finally, the Taylor expansion of the equity value smooth-pasting condition in x , equation (1.3.22g), gives the correction to the default boundary as:

$$\sqrt{\delta} x_{B,1}^y \frac{\partial^2 E_0^y}{\partial x^2} \left(x_{B,0}^y \right) = -\sqrt{\delta} \frac{\partial E_1^y}{\partial x} \left(x_{B,0}^y \right). \quad (1.4.20)$$

Equations (1.4.16a)-(1.4.16c), (1.4.19a)-(1.4.19c), and (1.4.20) form a system of equations which can be jointly solved to determine the first-order corrections for the debt value, the equity value, and the default boundary. While I am unable to give closed-form solutions to these corrections, it is now very straightforward to solve for them numerically since the differential equations are one-dimensional and the problems have fixed rather than free boundaries. The MATLAB function *bvp4c*, which implements a finite elements scheme, is used to solve the boundary value problems while searching over $\sqrt{\delta} x_{B,1}^y$ to satisfy equation (1.4.20). Note that I have not used condition (1.3.22h) at all; however, it is easy to show that the approximation $E_0(x, y) + \tilde{E}_1^\delta(x, y)$ is smooth across the boundary $x_{B,0}(y)$ as required.

To summarize, by making an assumption on the rate of mean-reversion of the process for volatility, this method reduces the solution of two-dimensional partial differential equation problems, in which one has a free boundary, to a recursive sequence of one-dimensional problems. The principal order terms simply reflect the contingent claims valuations and default boundary in the limiting case where volatility is fixed at its current level. The first-order correction terms are solved as a system of equations in which the contingent claims corrections are the solutions to one-dimensional, fixed boundary problems and the source terms are functions of the comparative statics or Greeks of the principal order terms. This method is tractable and will provide intuitive expressions demonstrating the effects of stochastic volatility on debt and equity valuation.

Moreover, a significant advantage of the methodology is that it reduces the number of parameters which need to be calibrated to generate quantitative results. Given the parametric assumption on the market price of variance risk, the only further calibration required beyond the baseline constant volatility model is that of two constants: $A^\delta = \sqrt{\delta} \nu_Y \Gamma_0$ and $B^\delta = \sqrt{\delta} \nu_Y \rho_Y$. Essentially, the variance risk premium needs to be calibrated as well as a constant which has the same sign as the correlation between volatility shocks and productivity shocks. The rate of mean-reversion κ_Y and the long-run mean of variance θ_Y do not appear at all.²⁷ This is especially useful in a structural model of credit incorporating stochastic volatility, as there are not good empirical estimates for the structural parameters of the asset volatility process.

Finally, I believe that this methodology is particularly well suited for use in a variety of other economic

²⁷This is because the operator \mathcal{M}_2 does not appear in the derivation of either the principal order term or the first-order correction term.

settings. Indeed, the approach offers a general framework for introducing additional state variables into either deterministic or stochastic dynamic models, including but certainly not limited to volatility. First, one constructs a baseline framework in which this additional state variable is a fixed parameter and then calculates the appropriate comparative statics. If one is able to make limiting assumptions on the dynamics of this additional state variable, such as slow, small, or fast, then approximate solutions to the full model can be derived by calculating correction terms as the solutions to differential equations whose source terms are functions of the comparative statics. For example, in the context of credit modeling, one could use this approach to develop a model which includes both stochastic volatility *and* stochastic interest rates.

1.4.3 Cumulative Default Probabilities

Given that I would like to study the credit spread puzzle, it is crucial that I am able to calculate cumulative default probabilities as well as valuations. As Huang and Huang (2003) point out, it is this computation in particular which has been one of the primary stumbling blocks in the construction of structural credit models which include stochastic volatility. My methodology, however, provides an effective means of overcoming this hurdle. The key, just as with the contingent claims valuations, is to find a suitable partial differential equations characterization of the probability and then utilize a perturbation to simplify the solution of the problem. Letting $u(l, x, y)$ denote the cumulative *survival* probability within l years, by the backwards Kolmogorov equation:

$$\left(-\frac{\partial}{\partial l} + \mathcal{L}_0 + \sqrt{\delta}\nu_Y\rho yx\frac{\partial^2}{\partial x\partial y} + \delta\mathcal{M}_2^y\right)u = 0 \quad (1.4.21a)$$

$$u(l, x_B(y), y) = 0 \quad (1.4.21b)$$

$$\lim_{x \rightarrow \infty} u(l, x, y) = 1 \quad (1.4.21c)$$

$$u(0, x, y) = 1, \quad (1.4.21d)$$

where $x_B(y)$ is the optimal default boundary from the equityholders' problem.²⁸ Importantly, note that the cumulative survival probabilities are calculated with respect to the dynamics of X_t and Y_t under the physical measure rather than the risk-neutral measure. Now expand this survival probability in powers of $\sqrt{\delta}$:

$$u(l, x, y) = u_0^y(l, x) + \sqrt{\delta}u_1^y(l, x) + \delta u_2^y(l, x) + \dots, \quad (1.4.22)$$

²⁸Recall that \mathcal{L}_0 is the Black-Scholes operator with a riskfree rate set equal to zero. Cumulative default probabilities are simply given by $1 - u(l, x, y)$.

substitute into equations (1.4.21a)-(1.4.21d), and expand the boundary conditions using Taylor expansions. Once again, the principal order term reflects the cumulative survival probability in the limiting case where volatility is fixed at the current level. I do not reproduce the expression here, but it can be looked up in any standard textbook treatment on the hitting times of geometric Brownian motion.

The correction term is an inhomogeneous partial differential equation with y as a parameter:

$$\left(-\frac{\partial}{\partial l} + \mathcal{L}_0\right) \sqrt{\delta} u_1^y = -B^\delta y x \frac{\partial^2 u_0^y}{\partial x \partial y} \quad (1.4.23a)$$

$$\sqrt{\delta} u_1^y(l, x_{B,0}^y) = -\sqrt{\delta} x_{B,1}^y \frac{\partial u_0^y}{\partial x}(l, x_{B,0}^y) \quad (1.4.23b)$$

$$\lim_{x \rightarrow \infty} \sqrt{\delta} u_1^y(l, x, y) = 0 \quad (1.4.23c)$$

$$\sqrt{\delta} u_1^y(0, x, y) = 0. \quad (1.4.23d)$$

As expected, the source term is a function of a comparative statics of the survival probability in the constant volatility case; however, contrary to the expressions above, the vega of the principal order term and the variance risk premium do not appear in equation (1.4.23a). The correction term can be calculated in MATLAB using the function *pdepe*, which implements a finite-difference scheme.

1.4.4 Debt Valuation under Exogenous Default

Finally, I will want to compare my results to a model in which the default trigger is specified exogenously, rather than determined endogenously. Let this exogenous boundary be given by \bar{x}_B . A perturbation approach can be used in a similar fashion as above to determine the approximate value of debt in the stochastic volatility model. As should be familiar by now, the principal order term reflects the constant volatility case and is simply given by equation (1.4.11) with $x_{B,0}^y$ replaced by \bar{x}_B . It turns out that in the case of $\rho_Y = 0$, a relatively simple explicit expression can be derived for the first-order correction term.

Theorem 1.4.1 *If the default boundary is set exogenously at \bar{x}_B and the correlation between productivity shocks and volatility shocks is equal to zero, then the first-order correction term in the asymptotic expansion of newly issued debt is given by*

$$\sqrt{\delta} \tilde{d}_1^y(x) = \frac{\ln(x/\bar{x}_B)}{2(g + \gamma_1 y - \frac{1}{2}y)} A^\delta y \left[\frac{\partial \tilde{d}_0^y}{\partial y} + \frac{1}{2(g + \gamma_1 y - \frac{1}{2}y)^2} y x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2} \right], \quad (1.4.24)$$

where γ_1 is the negative root of equation (1.4.12).

Proof. See Appendix A.1.3. ■

Thus, the first-order correction term involves both the vega and gamma of the value of newly issued debt in the constant volatility model. These are provided explicitly in Appendix A.1.3. In fact, an explicit expression can be computed for the case of $\rho_Y \neq 0$ as well, but it and its derivation are particularly cumbersome. Moreover, it is not needed in subsequent work and I therefore omit it. Note, however, that the method of derivation is very similar to that described in the proof of the expression above.

While the primary focus of this paper is on debt which allows for endogenous default by equityholders, this result is interesting and useful in its own right from an asset pricing perspective. There are forms of debt in which the exogenous trigger is in fact more appropriate. For instance, the debt may have certain covenants which enforce a zero net worth requirement as discussed in Leland (1994b).²⁹ Alternatively, default may occur once there is insufficient cash flow to meet the debt servicing obligations and new equity cannot be raised to make up the shortfall, a scenario which likely most accurately describes municipal debt. The expression above, along with the principal order term, allow for a closed-form solution to the pricing of such debt in a setting of stochastic volatility.

1.5 Equity Valuation

I begin my analysis by examining the model's qualitative predictions for equity pricing. Applying Ito's formula, substituting in equation (1.3.22a), and taking expectations shows that the required return of a firm's equity at time t under the physical measure is given by

$$\frac{(1 - \phi)(x - C) + \tilde{d}(x, y) - p + \mathbb{E}_t[dE^\delta]}{E^\delta} - r = \pi_X(y)\beta_X + \sqrt{\delta y}\nu_Y\Gamma(y)\beta_Y, \quad (1.5.1)$$

where the market and variance betas β_X, β_Y are respectively defined in equations (1.3.9) and (1.3.10). The excess expected return of equity over the riskfree rate is the sum of a risk premium due to its exposure to productivity risk and a risk premium due to its exposure to volatility risk. The exposures are priced according to the asset risk premium and variance risk premium, respectively. The first term is standard, while the second is novel. Note that the qualitative nature of expected returns will largely be driven by the comparative statics of the principal order terms. The model potentially resolves a number of empirical puzzles which have been documented in cross-sectional equity pricing and generates testable predictions.

²⁹Leland specifically discusses the example of a contractual arrangement in which the firm has access to a continuously renewable line of credit. Debt is rolled over at each instant at the fixed interest rate if and only if the firm's asset value is sufficient to cover the loan's principal. If not, default occurs.

1.5.1 Financial Distress and Momentum

Dichev (1998) and Campbell et al. (2008) find that the equity of financially distressed firms, i.e. those close to default, have lower returns on average than healthy firms, despite having higher market betas. That firms close to default have higher market risk is not surprising; indeed, this is the usual leverage effect. Then, according to a model in which market fluctuations are the only source of priced risk, financially distressed firms should, on average, earn higher returns. Hence, there is a puzzle.

If stochastic volatility is an additional source priced source of risk, though, then a firm's exposure to volatility risk also affects expected returns. An increase in volatility has both a positive and negative impact on the equity value. First, an increase in the volatility raises the value of the equityholders' default option, thereby positively impacting the equity valuation. Intuitively, since the equityholder's downside is capped by limited liability, while the upside potential is unlimited, an increase in uncertainty is beneficial.

However, the effect of higher volatility on rollover risk presents an offsetting negative channel. Debt is a concave function of productivity, reflecting the fact that the probability of default is most sensitive to movements in productivity near the default barrier. Since the flow dividend accruing to equityholders depends on the market value of newly issued debt, it too is concave in firm productivity. This concavity implies that an increase in volatility decreases the expected present value of future dividends.

The equity value of a financially distressed firm is largely comprised of its default option. For such firms, the first effect dominates, indicating that the value of equity rises with volatility. That is, a financially distressed firm's equity hedges against volatility risk in the market, or $\partial (\log E^\delta) / \partial y > 0$. If the variance risk premium is negative, reflecting a view by investors that persistent increases in volatility represent deteriorations in the investment opportunity set, then investors will be willing to accept a lower return on such equities due to their hedge value. On the other hand, the default option is less important in the valuation of healthy firms. For these firms, it is the exacerbating effect of volatility on rollover risk which dominates and equity values fall when volatility increases. Given this exposure to volatility risk, the required return on a sufficiently healthy firm will be higher than in a model with constant volatility.

Panel A of Figure 1.5.1 illustrates these points. It shows the variance betas of mature firms with the same book leverage as a function of financial distress and the average maturity of their debt. As expected, averaging across maturities, the variance betas of financially distressed firms are higher than those of healthy firms. Thus, the model offers a mechanism potentially resolving the financial distress puzzle. Furthermore, the figure generates new, testable predictions. First, note that variance betas decline as the average debt maturity of the firm decreases. The shorter the maturity structure, the greater the fraction of total principal which has to be rolled over at any point in time. Consequently, as the maturity structure shortens, rollover

risk becomes a greater component of the present value of future dividends accruing to equityholders, which in turn implies that firms with shorter maturity debt are more adversely impacted by increases in volatility. In terms of returns, by equation (1.5.1), this pattern of betas indicates that once market risk has been controlled for, firms with shorter maturity debt should have higher average returns than firms with longer maturity debt.

Panel B of Figure 1.5.1 also shows that for firms with shorter debt maturity, there is a clear hump-shaped relationship between financial distress and the equity vega.³⁰ That is, a decrease in the probability of default will increase the variance beta (towards zero) if the firm is sufficiently healthy. This is due to the fact that the effect of volatility on rollover risk is ameliorated as the health of the firm continues to increase. To understand this, note that the value of newly issued debt asymptotes to $(c + mp) / (r + m)$ as $x \rightarrow \infty$. In other words, debt becomes approximately flat as a function of productivity when the firm is very healthy. But in the region in which the debt valuation is flat, the firm faces little rollover risk and thus increases in volatility have little impact on the expected present value of future dividends. Equation (1.5.1) therefore indicates that variance risk premia should be maximized at intermediate probabilities of default. In fact, Garlappi and Yan (2011) find evidence supportive of this in their empirical work, documenting that average equity returns are hump-shaped as a function of the KMV distance to default.

Finally, consider forming a portfolio of financially distressed firms which is long recent winners and short recent losers. It is likely that, on average, the financial health of the winners has improved while that of the losers has deteriorated further. Then, by Figure 1.5.1, the variance betas of the losers should, on average, be higher than that of the winners. Thus, the winners should require higher variance risk premia than the losers, indicating that the portfolio should earn a positive CAPM alpha. This is consistent with the empirical evidence provided by Avramov et al. (2007) that the profits of momentum strategies are highly concentrated among a small subset of firms with low credit ratings. In fact, another empirical prediction of the model is that the momentum relation should reverse among healthy firms with a short debt maturity structure due to the hump-shaped nature of variance betas. Specifically, a portfolio which is long losers and short winners should earn a positive CAPM alpha.

1.5.2 Book-to-Market Effects

The implications of the model for the value premium puzzle are somewhat more subtle. The first channel I highlight is intimately related to the effects of financial distress discussed above. Consider two financially distressed, mature firms with the same book leverage and maturity structure. Suppose, though, that one firm

³⁰There is a hump-shaped relationship at longer maturities as well, but it is harder to detect.

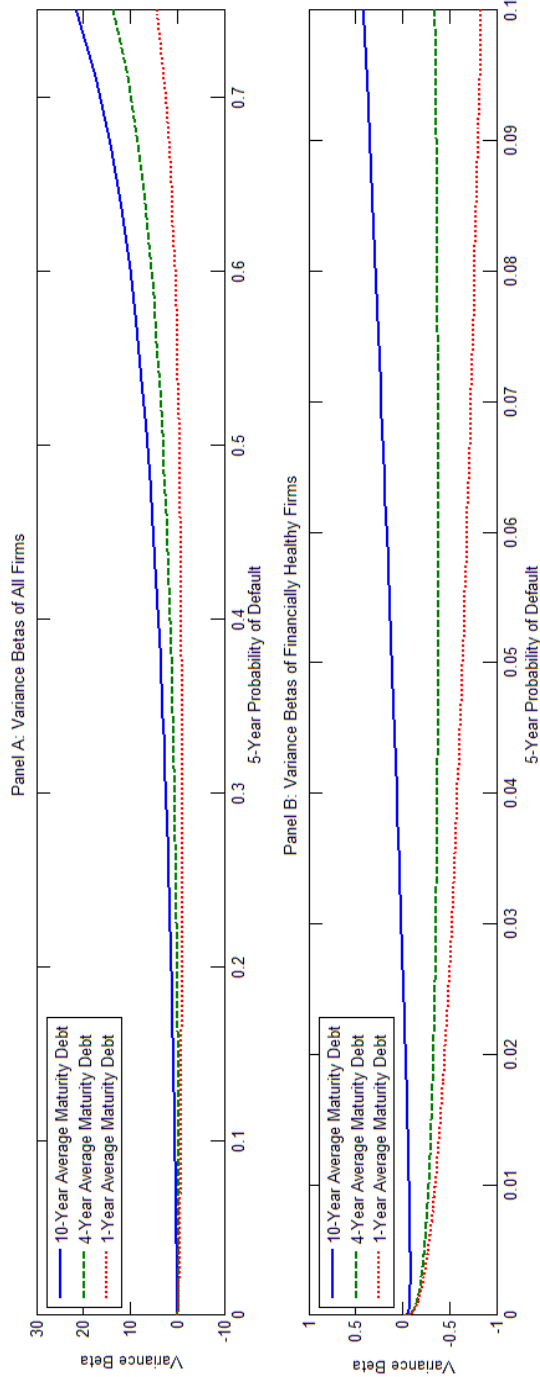


Figure 1.5.1: Variance Betas and Financial Distress

Note: This figure illustrates how equity variance betas vary with the average debt maturity of the firm and financial health of the firm, as measured by the 5-year cumulative probability of default. Firms are mature. The riskless rate, corporate tax rate, risk-neutral rate of productivity growth, and asset risk premium are set according to Table 1.6.1. Total principal equals 43.3 and the coupon rate equals 8.168%. Volatility of the productivity process equals 22%. Panel A shows the variance risk premia for all firms, while Panel B zooms in on financially healthy firms.

has a higher productivity, and thus lower book-to-market ratio, than the other. This is similar to Gomes, Kogan, and Zhang (2003) in that differences in the book-to-market ratio can be driven by cross-sectional variation in productivity across firms. Note then, however, that a *growth premium* actually emerges as a consequence of the explanation above for the financial distress puzzle. On the other hand, if the two firms are both healthy and one has a higher productivity, then the figure above shows that it is the firm with the higher book-to-market ratio that has the greater risk-adjusted return if the average maturity of debt is low. In this fashion, the model is capable of generating both a value premium and a growth premium depending on the level of financial distress.

Alternatively, two firms in the model with the same book leverage, productivity, and maturity structure can differ in their book-to-market ratios if they vary in the ratio of growth options to book value. As with financially distressed firms, much of the value of the young/growth firm in the model is comprised of its option to expand its installed capital. By the same logic as before, since volatility increases raise the value of this embedded option, growth firms should hedge against volatility risk in the market, more so than mature firms of similar default risk which have already exercised their growth options. If variance risk carries a negative price in the market, then investors will demand a higher variance risk premium to hold value stocks than growth stocks, all else equal.

1.6 Debt Valuation

I now move beyond qualitative considerations and turn to my primary quantitative analysis of the model's implications for debt pricing. I confine myself to analyzing the debt of mature firms. In order to study the debt valuations and default probabilities generated by the model, I must first articulate a suitable calibration of the key parameters.

1.6.1 Calibration

Given that my interest is in studying the credit spread puzzle, I will use a calibration procedure consistent with those previous studies which have demonstrated the puzzle, with a few notable exceptions. Most importantly, I will not calibrate the asset volatility to force the model to match historical default probabilities, but rather set the asset volatility according to *model-free* empirical estimates and then ask if the model is able to *jointly* generate reasonable credit spreads and default rates by credit rating.³¹ This is in marked contrast to Huang and Huang (2003) who require the models they analyze to match historical default

³¹Recall that asset volatility is the same as the volatility of productivity.

Table 1.6.1: Calibration of Model Parameters

r	ϕ	g	π_X
.08	.15	.02	.04

Note: This table provides the calibration values for those parameters which do not vary by credit rating: interest rate r , tax rate ϕ , risk-neutral productivity growth g , and asset risk premium π_X .

frequencies. The problem with the Huang and Huang approach, as I will demonstrate, is that the implied asset volatilities can be unreasonably high given the empirical estimates, especially at short maturities. For computational reasons, I also do not set the bankruptcy cost ξ to match recovery ratios by rating category, as does Huang and Huang (2003), but instead set this parameter directly and verify ex-post that recovery rates are approximately equal to the historical average of 51%.³² This approach is consistent with Leland (2004).

Table 1.6.1 summarizes those parameters which do not vary by credit rating. These include the interest rate, the corporate tax rate, the rate of productivity growth under the risk-neutral measure, and the asset risk premium. The riskfree rate is set equal to 8% as in Huang and Huang (2003) and Leland (2004), the historical average of Treasury rates between 1973-1998. The tax rate is equal to 15% as in Leland (2004), reflecting the corporate tax rate offset by the personal tax advantage of equity returns.³³ I set the rate of productivity growth equal to 2%. This indicates that the expected return on the value of assets in place is equal to 2% by equation (1.3.13), reflecting a payout rate of 6%. Finally, I set the asset risk premium equal to 4% such that the asset beta is equal to approximately 0.6 for all credit ratings. This is once again consistent with Leland (2004) and is slightly less than the asset risk premia in Huang and Huang (2003). Note that the asset risk premium does not affect the pricing of corporate debt and only impacts cumulative default probabilities.

Table 1.6.2 details those parameters of the model which do vary by credit rating, including leverage ratios, bankruptcy costs, average maturity of debt, and finally asset volatility. Target leverage ratios are from Standard & Poors (1999) and are consistent with both Huang and Huang (2003) and Leland (2004). As discussed previously, bankruptcy costs are set such that ex post recovery rates are approximately equal to 51%. Fractional costs of $\xi = 30\%$ works well for all credit ratings except for Caa, which has a slightly higher cost of $\xi = 35\%$. Average maturities and average asset volatilities are from Schaefer and Strebulaev (2008).

³²Specifically, Huang and Huang (2003) set bankruptcy costs by credit rating so that the models generate recovery rates of exactly 51.31%.

³³Specifically, as shown in Leland (2004), given a corporate tax rate of 35%, a personal tax on bond income of 40%, and a tax rate on stock returns of 20%, the effective tax advantage of debt can be calculated as $1 - (1 - 0.35)(1 - 0.20)/(1 - 0.40) = 0.133$.

Table 1.6.2: Calibration of Model Parameters by Credit Rating

Parameter	Credit Rating						
	Aaa	Aa	A	Baa	Ba	B	Caa
Target Leverage	13.1%	21.1%	32.0%	43.3%	53.5%	65.7%	80.0%
Avg. Asset Vol.	22%	22%	22%	22%	23%	28%	28%
Bankruptcy Costs	30%	30%	30%	30%	30%	30%	30%
Avg. Maturity (yr.)	10.16	9.45	10.13	9.14	7.11	7.39	2.48

Note: This table provides calibration values for the target leverage ratio, average asset volatility, fractional bankruptcy costs, and average maturity of debt by credit rating.

Their estimation of the asset volatilities warrants further discussion. Importantly, these estimates are model free, to the extent discussed below, and are therefore not dependent on assuming a particular structural model of credit, which would make them ineligible for use. Specifically, the authors estimate asset volatility of a firm j at time t , denoted as $\sigma_{A_{j,t}}^2$, according to

$$\widehat{\sigma_{A_{j,t}}^2} = (1 - L_{j,t})^2 \sigma_{E_{j,t}}^2 + L_{j,t}^2 \sigma_{D_{j,t}}^2 + 2L_{j,t} (1 - L_{j,t}) \sigma_{ED_{j,t}}, \quad (1.6.1)$$

where $L_{j,t}$ is the market leverage of firm j at time t , $\sigma_{E_{j,t}}$ is the equity volatility at time t , $\sigma_{D_{j,t}}$ is the debt volatility at time t , and $\sigma_{ED_{j,t}}$ is the covariance between debt and equity returns at time t . The volatilities and covariances are calculated directly from the time series of equity and debt returns. While this estimation procedure is indeed ostensibly independent of a specific model, the authors do implicitly assume that movements in the asset value are the only source of fluctuations in the debt and equity values. This is, of course, not the case in a model with stochastic volatility, as movements in the volatility level also constitute a source of fluctuations in debt and equity values. *However*, given my assumption that movements in asset volatility are in fact *slow* in the manner described above, these estimates are accurate to principal order.³⁴

Finally, I initially set $\rho_Y = 0$; that is, productivity shocks and variance shocks are uncorrelated. The variance risk premium, i.e. the constant A^δ , is set to generate the target credit spread on 10-year Baa debt, but then the implied specification for the market price of variance risk is then held constant for all other credit ratings and at other maturities. I set the current productivity level $X_0 = 7.0588$ such that the current value of assets in place is equal to 100. For each credit rating, I set the total principal P equal to the target leverage ratio multiplied by 100 and then solve for the coupon such that newly issued debt, and the total current value of debt, is priced at par. This will imply that the credit spread can be calculated as the

³⁴ Given a regime of slowly mean-reverting asset volatility, the estimates of Schaefer and Strebulaev (2008) will slightly overestimate the true average asset volatilities.

coupon rate C/P .

1.6.2 Credit Spreads on Intermediate/Long Maturity Debt

I begin my quantitative analysis by examining the credit spreads of 10-year and 20-year maturity debt. The second column of Table 1.6.3 reports the target historical credit spreads for 10-year maturity debt by rating category. The targets for investment grade through speculative grade debt (Aaa-Baa) are from Duffee (1998) while the targets for speculative grade through junk debt (Ba-Caa) are from Caouette, Altman, and Narayanan (1998). The questions I seek to answer are twofold. Is it possible to match historical credit spreads on intermediate to long maturity debt with a reasonable variance risk premium and can a single specification for the market price of variance risk explain credit spreads across credit ratings and maturities?

First, consider the performance of the model for 10-year maturity debt in which volatility is constant. The results are reported in the third and fourth columns of Table 1.6.3. As is evident, I recover the credit puzzle in this baseline model, especially for investment grade and speculative debt. For all credit ratings between Aaa-Baa, the model is never able to account for more than 30% of the target credit spread, although the performance is increasing as the rating worsens. On the other hand, the table illustrates that there is significantly less of a credit puzzle for junk debt. The baseline model performs much better at these ratings, explaining approximately 61% of the historical B credit spread and 87% of the historical Caa credit spread. This observation highlights one of the key challenges that a model needs to overcome to fully resolve the credit spread puzzle. In other words, it is important not to create a credit spread puzzle in the other direction, whereby the new model is able to better explain the credit spreads of investment grade debt, but then *overpredicts* the credit spreads of junk debt.

This is essentially what happens in the model incorporating stochastic volatility but in which the default boundary is specified *exogenously*. As can be seen from columns 5 and 6 of the table, the model is now able to explain a substantially higher proportion of the target credit spread for investment grade and speculative debt than the baseline model. However, the model significantly overpredicts the credit spreads on junk debt. The credit spread on the B-rated debt is overpredicted by 30% and the credit spread on Caa-rated debt is substantially overpredicted by 70%. This is not a particularly compelling resolution of the credit spread puzzle.

Conversely, the model incorporating stochastic volatility in which the default boundary is determined *endogenously* performs much better. The model now only slightly overpredicts the spreads on junk debt. B-rated debt is overpredicted by only 8% and Caa-rated debt by only 15%, a substantial improvement. Not only that, but the endogenous default model outperforms the exogenous default model for investment grade

Table 1.6.3: Credit Spreads on 10-Year Maturity Debt (bps)

	Target	Baseline		Exog. Default		Endog. Default	
		Model	% Explained	Model	% Explained	Model	% Explained
Aaa	47	2	4.3%	14	29.8%	19	40.4%
Aa	69	6	8.7%	37	53.6%	46	66.7%
A	96	19	19.8%	84	87.5%	90	93.8%
Baa	150	43	28.7%	150	100.0%	150	100.0%
Ba	310	93	30.0%	260	83.87%	242	75.6%
B	470	286	60.9%	607	129.2%	508	108.1%
Caa	765	663	86.7%	1307	170.9%	879	114.9%

Note: This table provides target and model-generated 10-year credit spreads by ratings category. Calibration parameters are provided in Tables 1.6.1 and 1.6.2. There is zero correlation between productivity shocks and variance shocks. The variance risk premium is set to match the 10-year historical Baa credit spread. Historical target credit spreads for Aaa-Baa debt are from Duffee (1998) while those for lower ratings categories are from Caouette, Altman, and Narayanan (1998). The baseline model holds volatility constant. The exogenous default model incorporates stochastic volatility, but sets the default barrier to be equal to that of the baseline model. The endogenous default model is the full stochastic volatility model.

debt as well, especially Aaa an Aa debt. While the exogenous default model accounts for approximately 30% and 54% of the target credit spreads for Aaa and Aa debt respectively, the endogenous default model can explain 40% and 67%. The only rating category that the exogenous default model wins in is Ba-rated debt. It explains 84% of the target credit spread relative to 76% for the endogenous default model, a slight improvement.

The calibration yields a parameter of $A^\delta = -.2264$, which in turn implies a variance risk premium of -1.1% for an asset volatility of 22% and a premium of -1.77% for an asset volatility of 28%. This result for A^δ is significantly lower than most estimates reported in the empirical literature using market returns and market volatility. This is to be expected and, indeed, it would be worrisome if the calibrated parameter were equal to or higher than empirical estimates. The reasons are twofold. First, since I am operating under the assumption of slowly-moving asset volatility, the volatility of asset variance should be significantly lower than the volatility of market variance. Second, innovations in the asset variance of an individual firm are likely only partially correlated with innovations in aggregate market variance, indicating that only a fraction of the volatility in asset variance can be accounted for by exposure to market variance risk. In other words, the Wiener process $W_t^{(2)}$ is only partially correlated with the process driving movements in market variance, which I denote by $W_t^{(m)}$. Thus, the price of $W_t^{(2)}$ risk should be lower in magnitude than the price of $W_t^{(m)}$ risk. This too will lead to a lower A^δ . Both of these explanations are consistent with the results of Carr and Wu (2009) who find that the expected returns on variance swaps of individual equities is well-explained by the stocks' volatility betas.

Using this value of A^δ , I next look at the credit spreads of 20-year debt in Table 1.6.4. While I

Table 1.6.4: Credit Spreads on 20-Year Maturity Debt (bps)

	Target	Baseline		Exog. Default		Endog. Default	
		Model	% Explained	Model	% Explained	Model	% Explained
Aaa	59	3	5.1%	28	47.5%	35	59.3%
Aa	87	8	9.2%	63	72.4%	70	80.4%
A	117	22	18.8%	121	103.4%	131	112.0%
Baa	198	46	23.2%	202	102.0%	194	98.0%
Ba	N/A	91	N/A	326	N/A	282	N/A
B	N/A	258	N/A	692	N/A	513	N/A
Caa	N/A	535	N/A	1256	N/A	716	N/A

Note: This table provides target and model-generated 20-year credit spreads by ratings category. Calibration parameters are provided in Tables 1.6.1 and 1.6.2. There is zero correlation between productivity shocks and variance shocks. The variance risk premium is set to match the 10-year historical Baa credit spread. Historical target credit spreads for Aaa-Baa debt are from Duffee (1998) while those for lower ratings categories are from Caouette, Altman, and Narayanan (1998). The baseline model holds volatility constant. The exogenous default model incorporates stochastic volatility, but sets the default barrier to be equal to that of the baseline model. The endogenous default model is the full stochastic volatility model.

unfortunately do not have target credit spreads for junk debt at this maturity, it is clear that there once again exists a significant credit puzzle for investment grade debt. A model without stochastic volatility is unable to explain more than 25% of the historical credit spreads at any rating. Adding stochastic volatility greatly improves the pricing of investment grade debt, with the endogenous default model once again outperforming the exogenous default model for Aaa and Aa debt by a substantial amount. I also report the credit spreads for junk debt and it is apparent that the exogenous default model yet again produces significantly higher credit spreads at these ratings categories than a model with endogenous default.

Finally, given that I do not force the model to match historical default rates, it is important to see what cumulative default probabilities the model is actually generating. In particular, I need to make sure that I am not delivering higher credit spreads by simply overstating the credit risk. The targets are given by the average cumulative issuer-weighted global default rates from 1970-2007 as reported by Moody's. The model is quite successful at matching long-maturity historical default rates for A, Baa and B rated debt as Table 1.6.5 demonstrates. It underpredicts the default probabilities in the Ba category somewhat and substantially so in the Aaa and Aa categories. Note, though, that the model-generated default probabilities for Aaa and Aa debt are much closer to reported historical rates from 1983-2007.³⁵

³⁵Moody's reports a global default rate of only 0.19% for Aaa debt between the years 1983-2007 at all maturities of 8 years and above. The rates for Aa-rated debt are also substantially lower in this period than between 1970-2007.

Table 1.6.5: Cumulative Default Probabilities - Long Maturity

	10yr		15yr		20yr	
	Target	Model	Target	Model	Target	Model
Aaa	0.53%	0.01%	1.00%	0.08%	1.20%	0.22%
Aa	0.52%	0.15%	1.09%	0.58%	1.88%	1.14%
A	1.31%	1.08%	2.40%	2.54%	4.08%	3.91%
Baa	4.35%	4.16%	7.60%	7.13%	10.51%	9.41%
Ba	18.43%	13.40%	27.53%	18.40%	34.85%	21.73%
B	40.92%	38.33%	50.21%	45.34%	52.38%	49.75%

Note: This table reports historical and model-generated cumulative default probabilities of firms within 10, 15, and 20 years by ratings category. Target expected default frequencies are the average cumulative issuer-weighted global default rates from 1970-2007 as reported by Moody's. Calibration parameters are provided in Tables 1.6.1 and 1.6.2. There is zero correlation between productivity shocks and variance shocks.

1.6.3 Intuition

These quantitative results naturally lead to two essential questions. How does the inclusion of stochastic volatility increase credit spreads and why is endogenous default important in matching credit spreads across ratings categories? The answer to the first question is intuitive. The negative market price of volatility risk indicates that investors see an increase in volatility as a deterioration in the investment opportunity set. Therefore, investors require a premium in the form of higher expected returns to hold assets which do poorly when volatility increases. In general, debt is such an asset since an increase in volatility raises the probability of default which lowers the value of the claim. Consequently, the discount rates on debt should be higher in a structural credit model with stochastic volatility and a negative market price of volatility risk than in a specification in which volatility is constant. These higher discount rates then lead to lower debt prices and higher credit spreads.

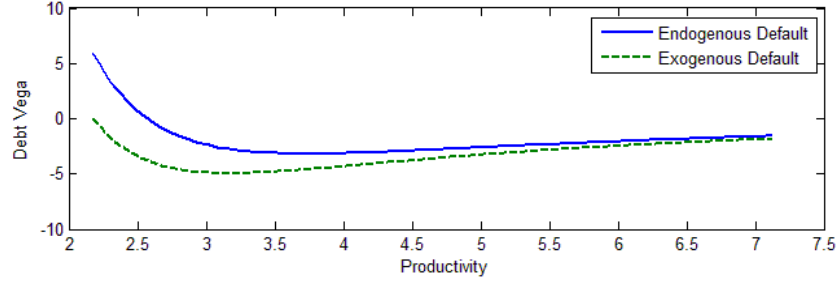
To see this explicitly, I apply Feynman-Kac to the boundary value problem defining the first-order correction for debt to derive a probabilistic interpretation. This gives:

$$\sqrt{\delta} \tilde{d}_1^y(x) = E_t^* \left[\int_t^{\tau^B(y)} e^{-(r+m)(s-t)} \left\{ - \left(A^\delta y \frac{\partial \tilde{d}_0^y}{\partial y} - B^\delta y \frac{\partial^2 \tilde{d}_0^y}{\partial x \partial y} \right) \right\} ds + e^{-(r+m)(\tau^B(y)-t)} \sqrt{\delta} x_{B,1}^y \left\{ m U' (X_{\tau^B(y)}) - \frac{\partial \tilde{d}_0^y}{\partial x} (X_{\tau^B(y)}) \right\} \right], \quad (1.6.2)$$

where the expectation is taken over the process $dX_s = gX_s + \sqrt{y}X_s dW_s^{(1)*}$ with $X_t = x$ and the stopping time defined by:

$$\tau^B(y) = \min \left\{ s : X_s = x_{B,0}^y \right\}.$$

In words, the correction term is an average discounted present value of comparative statics in the constant volatility model over all possible sample paths of productivity given an initial value of x , taking into account



Note: This figure illustrates how debt vegas vary with productivity in the constant volatility model under endogenous and exogenous default assumptions. Firms are mature. The riskless rate, corporate tax rate, risk-neutral rate of productivity growth, and asset risk premium are set according to Table 1.6.1. Total principal equals 43.3 and the coupon rate equals 8.168%. Current volatility of the productivity process equals 22%.

Figure 1.6.1: Debt Vegas: Endogenous and Exogenous Default

the stopping time and in which the process for productivity is a geometric Brownian motion with drift g and constant volatility \sqrt{y} . At the stopping time, the payoff is the correction to the debt value at the default boundary. The fact that the averaging holds volatility constant reflects the assumption of a slow-moving variance process. The first term in the expression captures the intuition described above. Recall that $A^\delta y$ is the variance risk premium and $\partial \tilde{d}_0^y / \partial y$ is the debt vega in the constant volatility model. So the correction term takes into account the extent to which debt in the constant volatility model covaries with variance, prices this risk according to the variance risk premium, and averages over all possible sample paths for productivity going forward. Since the vega should in general be negative for debt and the variance risk premium is negative, this constitutes a negative contribution to the correction term, raising credit spreads beyond the constant volatility baseline model.

The second term was not operative in the quantitative results previously since I set $\rho_Y = 0$, which implies $B^\delta = 0$. However, in a model with nonzero correlation, this term would capture the skewness effect of stochastic volatility. A negative correlation between productivity and volatility shocks means that bad times for the firm are more volatile times. This increases the credit risk of debt, i.e. the probability of default, which increases credit spreads. Technically, the cross partial $\partial^2 \tilde{d}_0^y / \partial x \partial y$ is generally positive and with negative correlation the contribution to the debt correction term is negative ($B^\delta < 0$). Note that while the previous effect was one of discount rates, the skewness effect is a statement about expected cash flows.

The mechanisms underlying the improved performance of the endogenous default model are more subtle. There are two driving forces. First, while the debt vega is indeed usually negative in the constant volatility model, it is actually positive when the firm is close to default if the barrier is chosen endogenously. This does not occur in the exogenous default model, as shown in Figure 1.6.1. When volatility increases, the option value of the equityholders increase and they respond, if able to, by postponing default. That is,

$x_{B,0}^y$ is a decreasing function of y . Increasing volatility therefore has two effects on the value of debt. The increased riskiness raises the probability of default directly, but the shifting boundary lowers it indirectly. Near the default barrier, the latter effect dominates and the increased volatility actually increases the value of debt. This can be seen mathematically. Differentiating equations (1.4.10a)-(1.4.10c) gives the problem to solve for the debt vega of the principal order term in the endogenous default model:

$$\mathcal{L}_{r+m}^y \frac{\partial \tilde{d}_0^y}{\partial y} + \frac{1}{2} x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2} = 0 \quad (1.6.3a)$$

$$\frac{\partial \tilde{d}_0^y}{\partial y} \left(x_{B,0}^y \right) = - \frac{\partial \tilde{d}_0^y}{\partial x} \left(x_{B,0}^y \right) \frac{dx_{B,0}^y}{dy} \quad (1.6.3b)$$

$$\lim_{x \rightarrow \infty} \frac{\partial \tilde{d}_0^y}{\partial y} (x) = 0 \quad (1.6.3c)$$

Since the value of debt is increasing at the default boundary and the boundary is decreasing with volatility, the debt vega at the boundary is positive. The result follows by continuity.

The second effect comes from the fact that in the probabilistic representation of the debt correction term, the payoff at the stopping time is positive. That is,

$$\sqrt{\delta} x_{B,1}^y \left\{ m U' \left(x_{B,0}^y \right) - \frac{\partial \tilde{d}_0^y}{\partial x} \left(x_{B,0}^y \right) \right\} > 0. \quad (1.6.4)$$

Intuitively, even the prospect of future movements in volatility increases the value of the equityholders' default option, leading to a lower default boundary than in the constant volatility case. Debtholders benefit from this because it lowers the probability of default. Technically, the term in brackets in equation (1.6.4) is negative. To see this, note that:

$$\begin{aligned} \frac{\partial \tilde{d}_0^y}{\partial x} \left(x_{B,0}^y \right) &= \frac{\partial V_0^y}{\partial x} \left(x_{B,0}^y \right) - \frac{\partial E_0^y}{\partial x} \left(x_{B,0}^y \right) \\ &= \frac{\partial V_0^y}{\partial x} \left(x_{B,0}^y \right) \\ &= U' \left(x_{B,0}^y \right) - \gamma_2 \left[\xi U \left(x_{B,0}^y \right) + \frac{\phi C}{r} \right] \\ &\geq U' \left(x_{B,0}^y \right), \end{aligned}$$

where V_0^y is the principal order term of total shareholder value as defined in Appendix A.1.2. The second inequality follows from smooth-pasting at the default boundary in the constant volatility model, the third inequality follows from differentiating the expression provided in Appendix A.1.2, and the final inequality follows from $\gamma_2 < 0$.

Crucially, note that these two effects are more significant in the pricing of junk debt than investment

Table 1.6.6: Credit Spreads on 4-Year Maturity Debt (bps)

	Target	Baseline		Exog. Default		Endog. Default	
		Model	% Explained	Model	% Explained	Model	% Explained
Aaa	46	0.5	1.1%	2	4.4%	2	4.4%
Aa	56	3	5.4%	13	23.2%	14	25.0%
A	87	12	13.8%	38	43.7%	42	48.3%
Baa	149	36	24.2%	91	61.1%	96	64.4%
Ba	310	94	30.3%	189	61.0%	193	62.3%
B	470	344	73.2%	541	115.1%	520	110.6%
Caa	N/A	1072	N/A	1570	N/A	1445	N/A

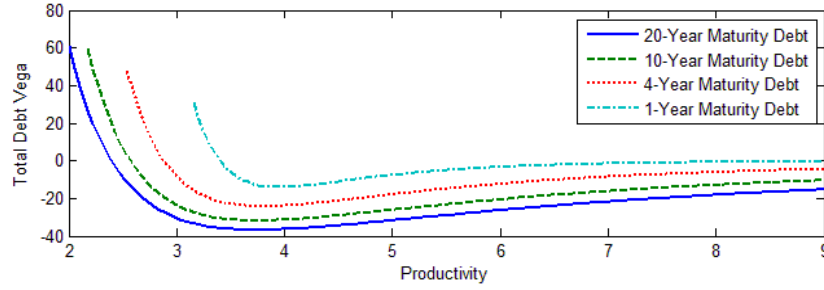
Note: This table provides target and model-generated 4-year credit spreads by ratings category. Calibration parameters are provided in Tables 1.6.1 and 1.6.2. There is zero correlation between productivity shocks and variance shocks. The variance risk premium is set to match the 10-year historical Baa credit spread. Historical target credit spreads for Aaa-Baa debt are from Duffee (1998) while those for lower ratings categories are from Caouette, Altman, and Narayanan (1998). The baseline model holds volatility constant. The exogenous default model incorporates stochastic volatility, but sets the default barrier to be equal to that of the baseline model. The endogenous default model is the full stochastic volatility model.

grade debt. When considering investment grade debt, default is far away and so most sample paths from the initial point will not hit the default boundary or the region near it for an extended period of time. As such, these two effects, which only occur near or at the default boundary, are heavily discounted among most sample paths. Thus they do not contribute much to the pricing of the correction term for investment grade debt, leading to similar pricing in both the endogenous and exogenous default models given a particular variance risk premium. On the other hand, it is exactly junk debt that is at risk of moving into the default region in the near future. Consequently, these effects are not discounted heavily in many of the sample paths from the initial point. Since these effects contribute positively to the value of debt, this indicates that the exogenous default model can significantly overpredict credit spreads.

It is now apparent why the endogenous default model is more successful. Since the variance risk premium is calibrated to the speculative grade Baa-rated credit spread, the implied variance risk premium is higher in the endogenous default model than the exogenous default model. However, the effects highlighted above are even stronger for junk debt than speculative grade debt. Therefore, the endogenous default model generates *lower* credit spreads at these ratings categories. On the other hand, since the effects are quite weak for investment grade debt and the implied variance risk premium is higher in the endogenous default model, it generates *higher* credit spreads at these categories as desired.

1.7 Short Maturity Debt and Extensions

Having quantitatively analyzed long maturity debt and discussed the mechanisms underlying the results, I now turn to a study of short maturity debt. As Table 1.6.6 illustrates, the credit spread puzzle is still



Note: This figure illustrates how total debt vegas in the constant volatility model vary with the average debt maturity and productivity of the firm. Firms are mature. The riskless rate, corporate tax rate, risk-neutral rate of productivity growth, and asset risk premium are set according to Table 1.6.1. Total principal equals 43.3 and the coupon rate equals 8.168%. Current volatility of the productivity process equals 22%.

Figure 1.7.1: Total Debt Vegas and Average Debt Maturity

present at short maturities and has familiar features. A model with constant volatility is unable to explain more than a third of the historical credit spread for Aaa-Ba debt, but accounts for a significantly greater fraction of junk credit spreads. Moreover, while the model with stochastic volatility and endogenous default certainly does improve on this baseline by a substantial amount, the performance is not as good as for long maturity debt. The model is never able to account for more than two-thirds of the historical credit spreads on investment and speculative grade debt. The performance for Aaa and Aa debt is particularly disappointing, with the model only accounting for 4.4% and 25.0% respectively of observed credit spreads.

One reason for this is that the effects of stochastic volatility are weaker at short maturities. In other words, the first-order correction term for total debt is smaller. This is because the vega of the principal order term is declining with maturity, as shown in Figure 1.7.1. Loosely stated, since productivity follows a diffusion, it can only move so far within a short period of time. Thus, increases in volatility do not significantly raise the riskiness of the firm. Since debt is not as sensitive to volatility fluctuations at short maturity, the discount rate correction in the stochastic volatility model is not as large. This is not the whole story, however. Examining Table 1.7.1 indicates that the model is underpredicting the cumulative default probabilities of short-maturity Aaa-Baa debt. That is, the model is not only underpredicting credits spreads, but also credit risk. A key question, therefore, is how well the model could match historical prices if it more accurately reflected empirical default frequencies.

The approach of Huang and Huang (2003) and other studies to this question has been to set the asset volatility to match this moment. As discussed, though, this method is somewhat unsatisfactory since the implied asset volatility is then significantly higher than model-free empirical estimates. A more appropriate approach is to ask whether the model as currently structured is missing some element of realism which, if included, could increase the credit risk of short maturity debt. One possibility would be to include negative

correlation between volatility shocks and productivity shocks, i.e. set $\rho_Y < 0$. This does not work as well as one would like though, since under the assumption of slow-moving volatility, the skewness effects are weak at especially short maturities.

Instead, I consider two extensions to the baseline model. In the first, I add a second, high-frequency factor in the volatility dynamics specification. In the second, I allow for rare disasters in the firm productivity process. In both cases, I will be able to extend the perturbation methodology to solve the model.

1.7.1 Multiscale Stochastic Volatility

Let \mathbf{W}_t now be a Wiener process or standard Brownian motion in three dimensions under \mathbb{P} . The dynamics of productivity and volatility are given by

$$dX_t/X_t = \mu dt + \sqrt{Y_t Z_t} dW_t^{(1)} \quad (1.7.1)$$

$$dY_t = \delta \kappa_Y (\theta_Y - Y_t) dt + \nu_Y \sqrt{\delta Y_t} dW_t^{(2)} \quad (1.7.2)$$

$$dZ_t = \frac{1}{\varepsilon} \kappa_Z (\theta_Z - Z_t) dt + \nu_Z \sqrt{\frac{1}{\varepsilon} Z_t} dW_t^{(3)}, \quad (1.7.3)$$

where both $\delta > 0$ and $\varepsilon > 0$ are small parameters and

$$\begin{pmatrix} W_t^{(1)} \\ W_t^{(2)} \\ W_t^{(3)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_Y & \sqrt{1 - \rho_Y^2} & 0 \\ \rho_Z & \rho_{YZ} & \sqrt{1 - \rho_Z^2 - \rho_{YZ}^2} \end{pmatrix} \mathbf{W}_t. \quad (1.7.4)$$

As in Calvet and Fisher (2004), the variance of productivity shocks is now a product of multiple factors. The first factor Y_t mean-reverts slowly just as before.³⁶ However, there now is an additional factor Z_t which mean-reverts quickly. The correlation between shocks to productivity and Y_t is given by ρ_Y , while the correlation between shocks to productivity and Z_t is given by ρ_{YZ} . The parameter ρ_{YZ} denotes the correlation between shocks to the high and low frequency components of volatility.

To price contingent claims, the dynamics under the risk neutral measure \mathbb{P}^* need to be specified. The dynamics of X_t and Y_t are given in equations (1.3.7) and (1.3.8). I assume that the dynamics of Z_t under the risk-neutral measure are the same as under the physical measure. This assumption is equivalent to saying that innovations in the high-frequency component to volatility are not priced. Long-run investors should not

³⁶Note that the parameter values κ_Y, θ_Y , and ν_Y may be different than before. However, under the assumption that $\delta > 0$ is small, the perturbation approach means that they will not need to be calibrated.

Table 1.7.1: Cumulative Default Probabilities - Short Maturity

	2yr		4yr		6yr	
	Target	Model	Target	Model	Target	Model
Aaa	0.00%	0.00%	0.03%	0.00%	0.17%	0.00%
Aa	0.02%	0.00%	0.11%	0.00%	0.26%	0.01%
A	0.09%	0.00%	0.34%	0.03%	0.61%	0.21%
Baa	0.48%	0.01%	1.36%	0.43%	2.32%	1.48%
Ba	3.02%	0.53%	7.65%	3.70%	11.77%	7.42%
B	10.20%	8.33%	20.33%	20.39%	28.74%	28.48%

Note: This table reports historical and model-generated cumulative default probabilities of firms within 2, 4, and 6 years by ratings category. Target expected default frequencies are the average cumulative issuer-weighted global default rates from 1970-2007 as reported by Moody's. Calibration parameters are provided in Tables 1.6.1 and 1.6.2. There is zero correlation between productivity shocks and variance shocks.

view highly transitory, independent innovations in volatility as deteriorations in the investment opportunity set. However, if shocks to high-frequency volatility are partially correlated with either innovations in asset productivity or innovations in low-frequency volatility, then they would carry a price. I ignore this effect since the goal at hand is to see how to increase the credit risk of short maturity debt, which the price of high-frequency volatility shocks would not affect.

The assumptions on capital structure and bankruptcy remain unchanged. The equity value $E(x, y, z)$ can again be characterized as the solution to a free boundary problem and the value of newly issued debt $\tilde{d}(x, y, z)$ as the solution to a PDE boundary value problem. To use the asymptotic expansion approach, equity values, debt values, and the default boundary are expanded in powers of both $\sqrt{\delta}$ and $\sqrt{\varepsilon}$:

$$E_{\delta, \varepsilon}(x, y, z) = E_0^{y, z}(x) + \sqrt{\delta}E_{1,0}^{y, z}(x) + \sqrt{\varepsilon}E_{0,1}^{y, z}(x) + \sqrt{\delta\varepsilon}E_{1,1}^{y, z}(x) + \dots \quad (1.7.5)$$

$$\tilde{d}_{\delta, \varepsilon}(x, y, z) = \tilde{d}_0^{y, z}(x) + \sqrt{\delta}\tilde{d}_{1,0}^{y, z}(x) + \sqrt{\varepsilon}\tilde{d}_{0,1}^{y, z}(x) + \sqrt{\delta\varepsilon}\tilde{d}_{1,1}^{y, z}(x) + \dots \quad (1.7.6)$$

$$x_B(y) = x_{B,0}^{y, z} + \sqrt{\delta}x_{B,1,0}^{y, z} + \sqrt{\varepsilon}x_{B,0,1}^{y, z} + \dots \quad (1.7.7)$$

A close adaptation of calculations in Fouque et al. (2003) based on both regular and singular perturbations derives the systems of equations which solve for the principal order terms and first-order correction terms. I summarize the results here.

Principal order terms once again reflect the equity/debt valuations and endogenous default boundary in the constant volatility model. However, the volatility plugged into these expressions is no longer \sqrt{y} , the current value of the slow-moving factor, but

$$\bar{\sigma}(y) = \sqrt{y} \int \sqrt{z} d\Lambda_Z(z), \quad (1.7.8)$$

where Λ_Z is the invariant distribution of Z_t as defined in Appendix A.3. That is, the volatility is set equal to the product of the slow-moving factor and the long-run average of the fast-moving factor. Note, therefore, that the principal order terms do not depend on the current level of Z_t .

First-order corrections are again the solutions to ordinary differential equations with a source term reflecting comparative statics of the constant volatility model. The slow-moving equity correction term is given by the same problem as before:

$$\mathcal{L}_r^{\bar{\sigma}^2(y)} \sqrt{\delta} E_{1,0}^{y,z} = \left(A^\delta y \frac{\partial E_0^{y,z}}{\partial y} - B^\delta y x \frac{\partial^2 E_0^{y,z}}{\partial x \partial y} - \sqrt{\delta} \tilde{d}_{1,0}^{y,z} \right) \quad \text{for } x > x_{B,0}^{y,z} \quad (1.7.9a)$$

$$\sqrt{\delta} E_{1,0}^{y,z} \left(x_{B,0}^{y,z} \right) = 0 \quad (1.7.9b)$$

$$\lim_{x \rightarrow \infty} \sqrt{\delta} E_{1,0}^{y,z}(x) = 0, \quad (1.7.9c)$$

except that the operator $\mathcal{L}_r^{\bar{\sigma}^2(y)}$ now reflects volatility $\bar{\sigma}(y)$ instead of \sqrt{y} . The fast-moving equity correction term is found according to

$$\mathcal{L}_r^{\bar{\sigma}^2(y)} \sqrt{\varepsilon} E_{0,1}^{y,z} = \left(C^\varepsilon \left[x^3 \frac{\partial^3 E_0^{y,z}}{\partial x^3} + 2x^2 \frac{\partial^2 E_0^{y,z}}{\partial x^2} \right] - \sqrt{\varepsilon} \tilde{d}_{0,1}^{y,z} \right), \text{ for } x > x_{B,0}^{y,z} \quad (1.7.10a)$$

$$\sqrt{\varepsilon} E_{0,1}^{y,z} \left(x_{B,0}^{y,z} \right) = 0 \quad (1.7.10b)$$

$$\lim_{x \rightarrow \infty} \sqrt{\varepsilon} E_{0,1}^{y,z}(x) = 0, \quad (1.7.10c)$$

where the constant C^ε has the opposite sign of ρ_Z .³⁷ Solving for the fast-moving correction term requires the gamma and speed of the constant volatility model. Both correction terms are independent of the current level of the fast-moving factor Z_t . The correction terms for debt are similarly found with Taylor expansions providing the boundary conditions for the differential equations. Finally, to complete the system of equations, the corrections to the default boundary must satisfy

$$\sqrt{\delta} x_{B,1,0}^{y,z} \frac{\partial^2 E_0^{y,z}}{\partial x^2} \left(x_{B,0}^{y,z} \right) = -\sqrt{\delta} \frac{\partial E_{1,0}^{y,z}}{\partial x} \left(x_{B,0}^{y,z} \right) \quad (1.7.11a)$$

$$\sqrt{\varepsilon} x_{B,0,1}^{y,z} \frac{\partial^2 E_0^{y,z}}{\partial x^2} \left(x_{B,0}^{y,z} \right) = -\sqrt{\varepsilon} \frac{\partial E_{0,1}^{y,z}}{\partial x} \left(x_{B,0}^{y,z} \right). \quad (1.7.11b)$$

Like the contingent claims, the endogenous default boundary is independent of the current Z_t . Cumulative survival probabilities can also be approximated using perturbation.

³⁷If innovations to the high-frequency component of volatility were priced, there would be an additional parameter in the coefficient multiplying gamma.

Table 1.7.2: Baa-Rated Credit Spreads with Multiscale Stochastic Volatility

	Target	$C^\varepsilon = 0$	$C^\varepsilon = .0005$	$C^\varepsilon = .0010$	$C^\varepsilon = .0015$	$C^\varepsilon = .0020$
20yr	195	194	200	207	212	218
10yr	150	150	155	161	168	174
4yr	149	95	102	108	115	120
1yr	N/A	28	35	41	46	52
3mo	N/A	3	5	5	7	8

Note: This table provides target and model-generated Baa-rated credit spreads at 20yr, 10yr, 4yr, 1yr, and 3mo maturities in a model with multiscale stochastic volatility. The parameter C^ε has the opposite sign of the correlation between shocks to the high-frequency volatility factor and productivity. Other calibration parameters are provided in Tables 1.6.1 and 1.6.2. $A^\delta = -.2264$ and $B^\delta = 0$. Historical target credit spreads are from Duffee (1998).

Quantitative Results and Discussion

I examine the credit spreads and default probabilities of Baa-rated debt as I vary the parameter C^ε . The choice of this rating category is for illustrative purposes only. The key intuitions and quantitative results carry over to other ratings. All other parameters in the model remain as before. In particular, I set $A^\delta = -.2264$ and $B^\delta = 0$. Table 1.7.2 shows that a increasing C^ε from zero to 0.002 raises credit spreads at all maturities. For instance, the spreads on 4-year maturity debt are raised from 95bps to 120bps, or from explaining 64.4% of the target spread to 80.5%.

Credit spreads rise because the model-generated cumulative default probabilities are increasing, as Table 1.7.3 illustrates. This is a reflection of the skewness effect described in Section 6. Recall that a positive C^ε indicates negative correlation between productivity shocks and shocks to fast-moving volatility. This means that bad times for the firm are also the most volatile times, which increases the probability of default.

The model achieves the stated goal of proportionally increasing credit spreads and default probabilities more at short maturities than at long maturities. As C^ε increases from 0 to 0.002, the credit spreads of 20-year and 10-year maturity debt increase by 11.8% and 16.0% respectively. The 4-year credit spread increases by 26.3%. Increases at shorter maturities, for which I do not have target credit spreads, are even more stark. The credit spreads on 1-year and 3-month debt increase by 85.7% and 166.7% respectively. Similarly, while the baseline model is only able to explain 31.6% of the probability of default within four years, setting $C^\varepsilon = .0015$ allows the model to match relatively well this target, while only overpredicting the probability of default within 20 and 10 years by 22.4% and 38.9% respectively.

The intuition behind this result is that periods of increased volatility correlated with negative market movements are highly transitory due the fast mean-reversion of the Z_t factor. Debt with long maturity will be averaging sample paths evolving over many years in which these skewness episodes will be brief and of small measure. Thus, they do not greatly impact the probability of default or valuation. Conversely,

Table 1.7.3: Baa-Rated Cumulative Default Probabilities with Multiscale Stochastic Volatility

	Target	$C^\varepsilon = 0$	$C^\varepsilon = .0005$	$C^\varepsilon = .0010$	$C^\varepsilon = .0015$	$C^\varepsilon = .0020$
20yr	10.51%	9.41%	10.64%	11.75%	12.86%	14.00%
10yr	4.35%	4.16%	5.14%	6.04%	6.95%	7.88%
4yr	1.36%	0.43%	0.77%	1.11%	1.46%	1.82%
2yr	0.48%	0.01%	0.05%	0.08%	0.12%	0.16%

Note: This table reports historical and model-generated cumulative default probabilities of Baa-rated firms within 20, 10, 4, and 2 years in a model with multiscale stochastic volatility. Target expected default frequencies are the average cumulative issuer-weighted global default rates from 1970-2007 as reported by Moody's. The parameter C^ε has the opposite sign of the correlation between shocks to the high-frequency volatility factor and productivity. Other calibration parameters are provided in Tables 1.6.1 and 1.6.2. $A^\delta = -.2264$ and $B^\delta = 0$.

such a skewness episode may coincide with the lifetime of especially short maturity debt despite the fast mean-reversion. This accounts for the large proportional increases observed at shorter maturities.

1.7.2 Rare Disasters

In my second extension to the baseline model, I allow for jumps in the firm productivity process:

$$dX_t/X_t = \mu_t dt + \sqrt{Y_t} dW_t^{(1)} + d \left(\sum_{i=1}^{N(t)} (Q_i - 1) \right) \quad (1.7.12)$$

$$dY_t = \delta \kappa_Y (\theta - Y_t) + \nu_Y \sqrt{\delta Y_t} dW_t^{(2)}, \quad (1.7.13)$$

where $N(t)$ is a Poisson process with rate $\lambda > 0$ and $\{Q_i\}$ is a sequence of independent, identically distributed random variables which take values between zero and one. Note that a jump in this model always corresponds to a negative event, although this could easily be modified. Jumps have been included in structural modeling previously by Hilberink and Rogers (2002) and Chen and Kou (2009). My contribution is to include jumps, stochastic volatility, and endogenous default in a unified model and to describe a tractable method for solving it.

I will assume that the jump risk premium is zero to maintain my focus on increasing the credit spreads of short maturity debt. Then, the risk-neutral dynamics are given by

$$dX_t/X_t = g dt + \sqrt{Y_t} dW_t^{(1)*} + d \left(\sum_{i=1}^{N(t)} (Q_i - 1) \right) \quad (1.7.14)$$

$$dY_t = \left(\kappa_Y (\theta_Y - Y_t) - \Gamma(Y_t) \nu_Y \sqrt{Y_t} \right) dt + \nu_Y \sqrt{Y_t} dW_t^{(2)*}. \quad (1.7.15)$$

A free boundary problem characterizes the equity valuation. The key innovation is that the equity value

must now satisfy an integro-partial differential equation, instead of a partial differential equation, to account for the jumps. The value of newly issued debt also satisfies an integro-partial differential equation.

The key assumption of the model is that the jumps are *rare*, i.e. $\lambda > 0$ is small. This allows me to utilize regular perturbation and expand the contingent claims and default boundary in powers of $\sqrt{\delta}$ and λ :

$$E_{\delta,\lambda}(x, y, z) = E_0^y(x) + \sqrt{\delta}E_{1,0}^y(x) + \lambda E_{0,1}^y(x) + \sqrt{\delta}\lambda E_{1,1}^y(x) + \dots \quad (1.7.16)$$

$$\tilde{d}_{\delta,\lambda}(x, y, z) = \tilde{d}_0^y(x) + \sqrt{\delta}\tilde{d}_{1,0}^y(x) + \lambda\tilde{d}_{0,1}^y(x) + \sqrt{\delta}\lambda\tilde{d}_{1,1}^y(x) + \dots \quad (1.7.17)$$

$$x_B(y) = x_{B,0}^y + \sqrt{\delta}x_{B,1,0}^y + \lambda x_{B,0,1}^y + \dots \quad (1.7.18)$$

As should be familiar by now, the principal order terms will reflect the model with constant volatility and no jumps. The first-order correction terms accounting for the slow-moving stochastic volatility will also be the same as before. The equity correction term accounting for the jumps is:

$$\mathcal{L}_r^y \lambda E_{0,1}^y = - \left(\lambda \int_0^1 [E_0^y(yx) - E_0^y(x)] f_Q(y) dy - \lambda \tilde{d}_{0,1}^{y,z} \right), \text{ for } x > x_{B,0}^{y,z} \quad (1.7.19)$$

$$\lambda E_{0,1}^{y,z}(x_{B,0}^y) = 0 \quad (1.7.20)$$

$$\lim_{x \rightarrow \infty} \lambda E_{0,1}^y(x) = 0, \quad (1.7.21)$$

where $f_Q(y)$ is the density of the jump size distribution. A similar correction term is derived for debt and, together with the corrections to the default boundary, form a system of equations.

Quantitative Results and Discussion

To calibrate the model, I set $\lambda = .001$ such that the probability of a disaster event in any given year is 1%. I assume the jump size distribution to be uniform over an interval $[Q_{\min}, Q_{\max}] \subset [0, 1]$. Note that this implies the firm will lose between Q_{\min} and Q_{\max} percent of its asset value in the event of a jump. Table 1.7.4 displays the credit spreads for Baa-rated debt for jump size intervals of $[.50, .75]$, $[.25, .75]$, and $[.25, .50]$. Clearly, as the magnitudes of the jumps increase, the credit spreads rise as well.

This model also achieves the goal of proportionally increasing short-maturity credit spreads more than long-maturity credit spreads. There is a 30.4% increase in the 20-year credit spread and 30.7% increase in the 10-year credit spread as one moves from a model with no jumps to a model with $\lambda = .001$ and $Q \sim U([.25, .50])$. Credit spreads on 4-year debt increase by 45.3%, while those on 1-year debt increase by 135.7%.

The reason behind this pattern is that credit risk is increased more proportionally at shorter maturities.

Table 1.7.4: Baa-Rated Credit Spreads with Multiscale Stochastic Volatility

	Target	$\lambda = 0$	$\lambda = .001$		
			[.50, .75]	[.25, .75]	[.25, .50]
20yr	195	194	216	236	253
10yr	150	150	165	180	196
4yr	149	95	109	124	138
1yr	N/A	28	37	50	66

Note: This table provides target and model-generated Baa-rated credit spreads at 20yr, 10yr, 4yr, 1yr maturities in a model with rare disasters. The parameter λ controls the arrival rate of the disasters and the intervals reflect the uniform distribution of the jump sizes. Other calibration parameters are provided in Tables 1.6.1 and 1.6.2. $A^\delta = -.2264$ and $B^\delta = 0$. Historical target credit spreads are from Duffee (1998).

The intuition is that the model allows for jumps to default. Since it is difficult for pure diffusions to reach the default barrier within a short period of time, the majority of the default probability for especially short maturity debt comes from this risk. This leads to large proportional increases in credit risk as one moves beyond a pure diffusion model. However, at longer maturities it is possible to both diffuse to default as well as jump to default. Therefore, the inclusion of jump risk does not proportionally increase credit risk as much.

1.8 Conclusion

This paper has constructed a real-options, term structure model of the firm incorporating stochastic volatility. Shocks to variance carry a negative price of risk in the market to reflect the view of long-term investors that a persistent increase in volatility constitutes a deterioration in the investment opportunity set. The stocks of financially distressed firms are long the default option and hedge against volatility risk in the market, thus requiring lower variance risk premia than healthy firms. Risk-adjusted momentum profits are concentrated among the stocks of low credit-rating firms and results from changing conditional variance betas. The model is capable of generating both a growth premium and a value premium among firms depending on the level of financial distress. A firm with growth options hedges volatility risk and, all else equal, requires smaller variance risk premia than firms without growth options, generating a value premium.

Firm debt is short the option to default such that if volatility is stochastic and shocks are negatively priced, discount rates are higher than in a model with constant volatility. This effect in turn lowers prices and raises credit spreads. However, it is the interaction of stochastic volatility and endogenous default which resolves the credit spread puzzle. Under a setting of endogenous default, extremely distressed junk debt hedges volatility risk in the market and the default threshold is lower at all levels of volatility than in the model with constant volatility. This prevents the model from overpredicting the credit spreads on junk

debt and also improves the performance of the model at higher ratings categories.

The paper has demonstrated a novel solution methodology from mathematical finance and physics which should have applicability in other areas of economics. Principal order terms in the asymptotic expansion reflect contingent claims valuations in the constant volatility setting, while correction terms are the solutions to ODEs involving key comparative statics of the constant volatility model. The approach yields tractability, reduces the number of parameters which need to be calibrated, and provides clean mathematical expressions to guide understanding of the mechanisms at work in the model.

Finally, the paper has considered an additional high-frequency volatility time scale and rare disasters as extensions to the primary model. These additional elements allow the model to increase credit risk at short maturities and improve the model's performance at matching empirical default frequencies and credit spreads at short maturities.

Chapter 2

How Do Foreclosures Exacerbate Housing Downturns?¹

2.1 Introduction

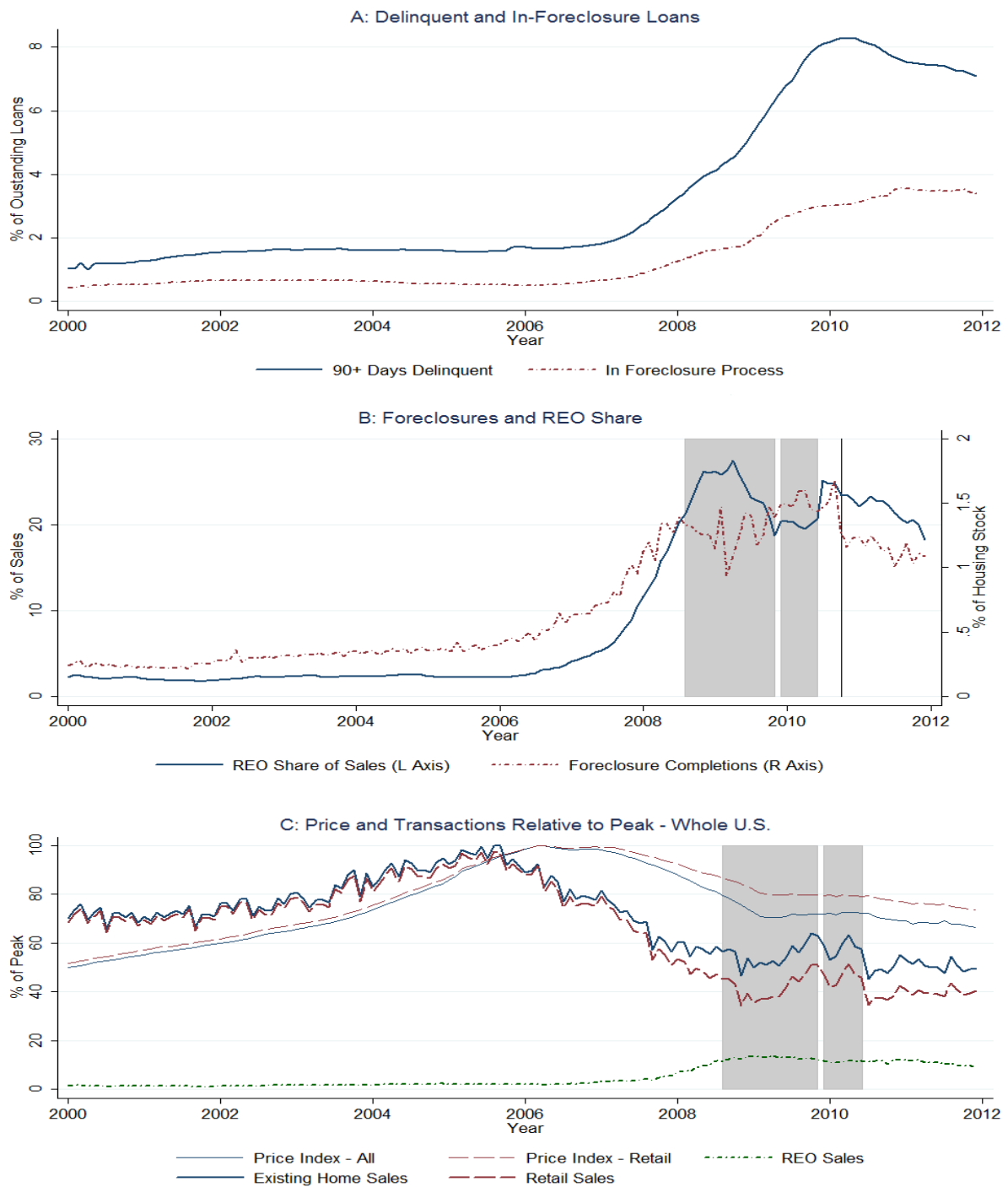
Foreclosures have been one of the dominant features of the recent housing market downturn. From 2006 through 2011, approximately 7.4 percent of the owner-occupied housing stock experienced a foreclosure.² Although the wave of foreclosures has subsided, foreclosures remain at elevated levels, and understanding the role of foreclosures in housing downturns remains an important part of reformulating housing policy going forward.

The behavior of the housing market concurrent with the wave of foreclosures is shown in Figure 2.1.1. Real Estate Owned (REO) sales – that is sales of foreclosed homes owned by banks and the GSEs – have made up between 20 and 30 percent of existing home sales nationally. Sales of existing homes fell 54.9 percent peak-to-trough; retail (non-foreclosure) volume fell 65.7 percent. Prices dropped considerably, with aggregate price indices plunging by a third and prices falling by a quarter for indices that exclude distressed sales. Time to sale and vacancy rates have also climbed, particularly in the retail market. Even with a slowdown in foreclosures due to lawsuits over fraudulent foreclosure practices, foreclosures have continued at a ferocious pace.

This paper presents a model in which foreclosures have important general equilibrium effects that can explain much of the recent behavior of housing markets, particularly in the hardest-hit areas. By raising the

¹Coauthored with Adam Guren.

²Data from CoreLogic. The data is described in Section 2.5 and Appendix B.4.



Note: All data is seasonally adjusted national-level data from CoreLogic as described in the data appendix. The grey bars in panels B and C show the periods in which the new homebuyer tax credit applied. The black line in panel B shows when foreclosures were stalled due to the exposure of fraudulent foreclosure practices by mortgage servicers. In panel C, all sales counts are unsmoothed and normalized by the total number of existing home sales at peak while each price index is normalized by its separate peak value.

Figure 2.1.1: The Role of Foreclosures in the Housing Downturn

number of sellers and reducing the number of buyers, by making buyers more choosy, and by changing the composition of houses that sell, foreclosures sales freeze up the market for retail sales and reduce both price and sales. Furthermore, the effects of foreclosures can be amplified considerably because price declines induce more default which creates further price declines, generating a feedback loop. A quantitative calibration suggests that these effects can be large: foreclosures exacerbate aggregate price declines by approximately 50 percent and retail price declines by 30 percent.

Despite the importance of foreclosures in the housing downturn, economists have not closely examined how the housing market equilibrates when there are a substantial number of distressed sales. A supply and demand framework, as employed by much of the financial literature on fire sales and illiquidity, can potentially explain declining prices and volumes with demand falling relative to supply but cannot speak to the freezing up of the retail market. Such models also assume that investors can adjust their positions continuously by transacting in a liquid market, yet housing is lumpy, illiquid, and expensive. A substantial literature has sought to adapt models to fit the peculiarities of the housing market and explain the positive correlation between volume and price. For instance, search frictions as in Wheaton (1990), Williams (1995), Krainer (2001), and Novy-Marx (2009), borrowing constraints as in Stein (1995), and nominal loss aversion as in Genesove and Mayer (2001) have been shown to play important roles in housing markets. Yet no paper has explicitly examined the role of distressed sales in a model tailored to housing.

To illustrate the mechanisms through which foreclosures affect the housing market, a simple model of the housing market with exogenous foreclosures is introduced. It adds two key ingredients to an otherwise-standard search-and-matching framework with stochastic moving shocks, random search, idiosyncratic house valuations, and Nash bargaining over price: REO sellers have higher holding costs and individuals who are foreclosed upon cannot immediately buy a new house. These two additions together dry up the market for normal sales, reduce volume and price, and imply that the market only gradually recovers from a wave of foreclosures. This occurs through three main effects. First, the presence of distressed sellers increases the outside option to transacting for buyers, who have an elevated probability of being matched with a distressed seller next period and consequently become more selective. This “choosy buyer effect” endogenizes the degree of substitutability between bank and retail sales. Second, because foreclosed individuals are locked out of the market, foreclosures reduce the likelihood that a seller will meet a buyer in the market through a “market tightness effect.” This effect emphasizes that foreclosures do not simply add supply to the market: a key feature of foreclosures is that they also reduce demand. Third, there is a mechanical “compositional effect” as the average sale looks more like a distressed sale.

The choosy buyer effect in particular is novel and formalizes folk wisdom in housing markets that foreclosures empower buyers and cause them to wait for a particularly favorable transaction. For instance,

The New York Times reported that “before the recession, people simply looked for a house to buy ... now they are on a quest for perfection at the perfect price,” with one real estate agent adding that “this is the fallout from all the foreclosures: buyers think that anyone who is selling must be desperate. They walk in with the bravado of, ‘The world’s coming to an end, and I want a perfect place.’”³ The Wall Street Journal provides similar anecdotal evidence, writing that price declines “have left many sellers unable or unwilling to lower their prices. Meanwhile, buyers remain gun shy about agreeing to any purchase without getting a deep discount. That dynamic has fueled buyers’ appetites for bank-owned foreclosures.”⁴ Although other papers such as Albrecht et al. (2007, 2010) and Duffie et al. (2007) have included seller heterogeneity in an asset market model, no paper that does so has generated a choosey buyer effect, which turns out to be important in explaining the disproportionate freezing up of the retail market.

To provide a more realistic treatment of the downturn, the basic model of the housing market is embedded in a richer model of mortgage default in which borrowers with negative equity may default on their mortgage or be locked into their current house despite a desire to move. This generates a new amplification channel: an initial shock that reduces prices puts some homeowners under water and triggers foreclosures, which cause more price declines and in turn further default. While reminiscent of the literature initiated by Kiyotaki and Moore (1997), the price declines here are caused by the general equilibrium effects of foreclosures. Lock-in of underwater homeowners also impacts market equilibrium by keeping potential buyers and sellers out of the market.

The richer model is used to quantitatively evaluate the extent to which foreclosures have exacerbated the ongoing housing bust. This quantitative analysis takes a two-pronged approach. First, we assess the strength of the amplification channel and its sensitivity to various parameters in the model. Second, we fit the model to data from the 100 largest MSAs to assess the empirical size of the amplification channel and test its implications across metropolitan areas. The model matches the data on the size of the price decline, the number of foreclosures, price declines in the retail market, and the REO share of sales. It also matches the heterogeneity in foreclosure discounts over the cycle found by Campbell et al. (2011). However, it falls short of explaining the full sales decline, suggesting that other forces have depressed transaction volume in the downturn. The quantitative analysis reveals that foreclosures exacerbate the aggregate price decline in the downturn by approximately 50 percent in the average MSA (or in other words account for a third of the decline) and exacerbate the price declines for retail sellers by over 30 percent.

Finally, we analyze the impact of the foreclosure crisis on welfare in our model and simulate three

³ “Housing Market Slows as Buyers Get Picky” June 16, 2010.

⁴ “Buyer’s Market? Stressed Sellers Say Not So Fast” April 25, 2011.

foreclosure-mitigating policies: slowing down foreclosures, refinancing mortgages at lower interest rates, and reducing principal. While we do not conduct a full normative analysis, the simulations of these policies highlight the trade-offs faced by policy makers.

The remainder of the paper is structured as follows. Before presenting the model, section 2.2 presents facts about the bust across metropolitan areas. To explain the data, the remainder of the paper develops a model of how foreclosures affect the housing market, first focusing on mechanisms and then on magnitudes. Section 2.3 introduces a model of exogenous defaults, and section 2.4 explores the intuitions and qualitative implications of the model. In section 2.5, the basic model is embedded in a more complete model in which negative equity is a necessary condition for default, which creates a new amplification mechanism in the form of a price-default spiral. The paper then turns to the magnitudes of the effects identified in sections 2.3-2.5. Section 2.6 calibrates the model and quantitatively analyzes the model's comparative statics and the strength of the price-default amplification channel. Section 2.7 takes the model to the national and cross-MSA data from the ongoing downturn. Section 2.8 considers welfare and foreclosure policy, and section 2.9 concludes.

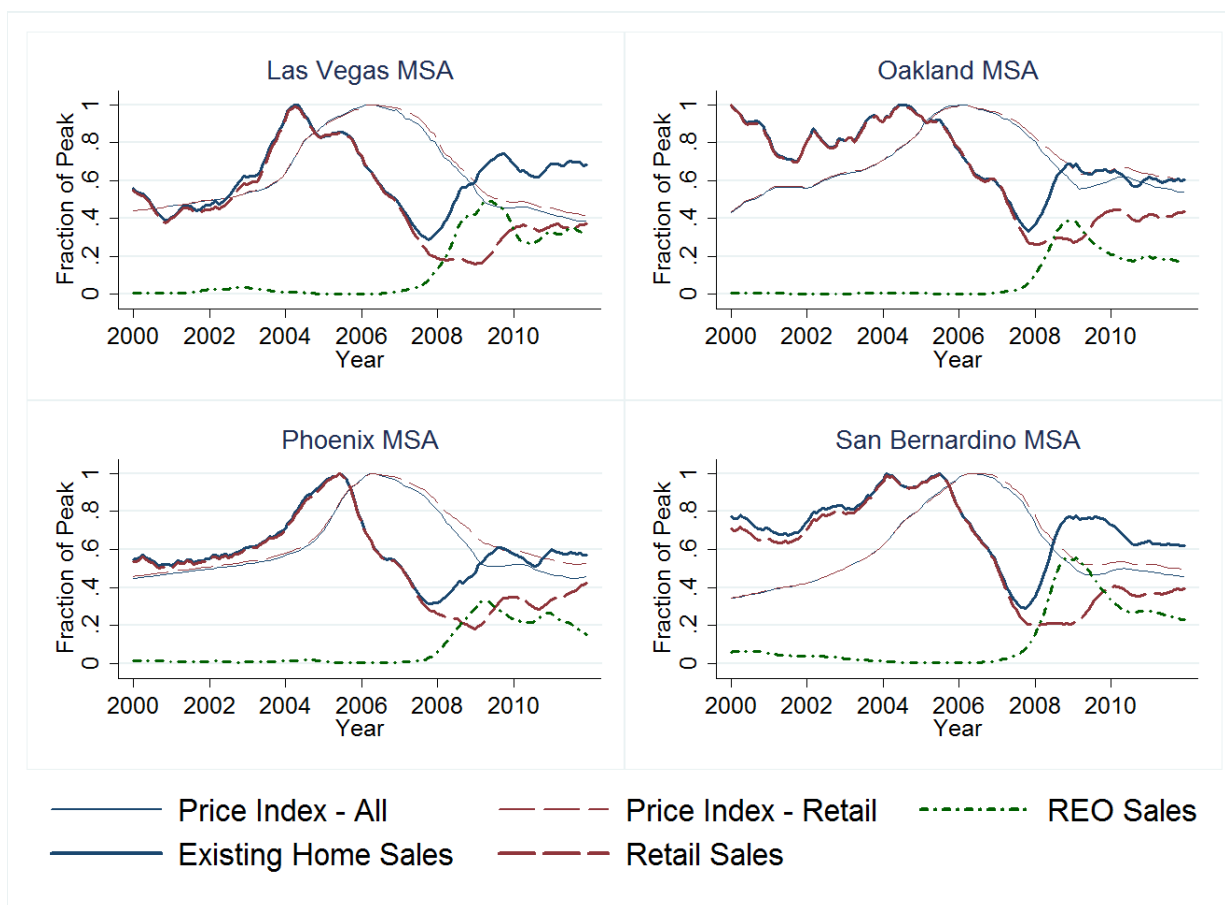
2.2 Empirical Facts

The national aggregate time series of price, volume, foreclosure, and REO share presented in Figure 2.1.1 mask substantial heterogeneity across metropolitan areas in the severity of the housing bust and wave of foreclosures. To illustrate this, Figure 2.2.1 shows price and volume time series for four of the hardest-hit metropolitan areas. In Las Vegas, for instance, prices fell 61.5 percent, retail sales fell 84.0 percent, and the REO share was as high as 76.4 percent. Figure 2.2.1 also illustrates how foreclosure sales substitute for retail sales: retail sales rise as REO sales recede and fall as REO sales surge.

To provide more systematic facts about the heterogeneity of the bust across MSAs, we use a proprietary data set provided to us by CoreLogic supplemented by data from the United States Census. CoreLogic provides monthly data for 2000-2011 for the nation as a whole and the 100 largest MSAs, from which we drop 3 MSAs because the full data are not available for these locations at the start of the crisis. The data set includes a house price index,⁵ a house price index for retail sales only,⁶ the number of completed foreclosure auctions, sales counts for REOs, new houses, existing houses (including short sales), and the estimates of quantiles of the LTV distribution described previously. These statistics are compiled by CoreLogic using

⁵The CoreLogic price index is a widely-used repeat sales index that has behaved similarly to other cited indices in that it fell by a third during the downturn. The S&P Case-Shiller index shows similar declines to the CoreLogic index. The FHFA expanded-data index, which includes FHFA data proprietary deeds data from other sources, fell 26.7 percent.

⁶Given the small number of distressed properties prior to the downturn, price indices for distressed properties are typically not estimated. The CoreLogic non-distressed price index drops REO sales and short sales from the database and re-estimates the price index using the same methodology.



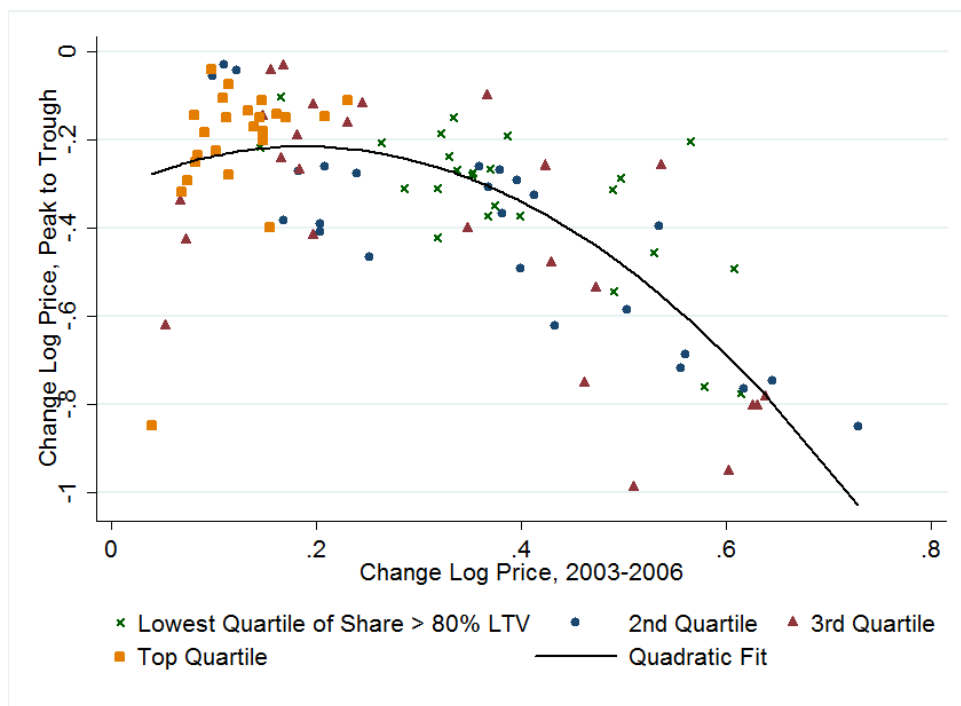
Note: All data is seasonally adjusted CBSA-level data from CoreLogic as described in the Appendix B.4. The sales lines are smoothed using a moving average. In panel C, all sales counts are normalized by the total number of existing home sales at peak while each price index is normalized by its separate peak value.

Figure 2.2.1: Price and Transaction Volume in Selected MSAs With High Levels of Foreclosures

public records. CoreLogic's data covers over 85 percent of transactions nationally. We seasonally adjust the CoreLogic data and smooth the sales count series using a moving average. A complete description of the data is in appendix B.4.

By far the best predictor of the size of the bust was the size of the preceding boom. Figure 2.2.2 plots the change in log price from 2003 to 2006 against the change in log price from each market's peak to its trough through 2011. There is a clear downward pattern, with the notable exception of a few outliers in the lower-left of the diagrams which correspond to metropolitan areas in southern Michigan which experienced a substantial bust without a large boom.

Figure 2.2.2 also reveals a more subtle fact in the data: places that had a larger boom had a more-than-proportionally larger bust. While a linear relationship between log boom size and log bust size has an r-squared of .44, adding a quadratic term that allows for larger busts in places with larger booms as illustrated in Figure 2.2.2 increases the r-squared to .57.



Note: Scatter plot of seasonally adjusted data from CoreLogic along with quadratic regression line. The data is fully described in Appendix B.4. Each data point is an MSA and is color coded to indicate in which quartile the MSA falls when MSAs are sorted by the share of homes with over 80% LTV in 2006. There are two main take-aways. First, there is a non-log-linear relationship between the size of the boom and the size of the bust, as the regression r -squared is .44 for a linear model and .57 for a quadratic model. Second, the color-coding of the points provides suggestive evidence that it is the combination of high LTV and a large bubble that is associated with a disproportionately large bust.

Figure 2.2.2: Boom vs. Bust Across MSAs

This paper argues that by exacerbating the downturn in the hardest-hit places, foreclosures can explain much of why the relationship between log boom size and log bust size is not linear. This explanation implies an additional reduced-form cross-sectional test: because default is closely connected to negative equity, a larger bust should occur in locations with the *combination* of a large bubble and a large fraction of houses with high loan balances – and thus close to default – prior to the bust. To provide suggestive evidence that this prediction is borne out in the data, the points in Figure 2.2.2 are color-coded by quartiles of share of houses in the MSA with over 80 percent LTV in 2006. While the highest measured LTVs came in places that did not have a bust – home values were not inflated in 2006, so the denominator was lowest in these locations – one can see that the majority of MSAs substantially below the quadratic trend line were in the upper end of the LTV distribution.

To investigate whether the interaction of high LTV and a big bust combined is correlated with a deep

Table 2.2.1: MSA Summary Statistics

	Unweighted Mean	SD	Min	Max	N
Max $\Delta \log (P)$	-.3398355	.2292549	-.9895244	-.0286884	97
Max $\Delta \log (P_{\text{Retail}})$	-.2784659	.1929688	-.9229212	-.0388126	97
Max $\Delta \log (\text{Sales}_{\text{Existing}})$	-.9354168	.2684873	-2.53161	-.4671416	97
Max $\Delta \log (\text{Sales}_{\text{Retail}})$	-1.174493	.3181327	-2.736871	-.5613699	97
$\frac{\text{Sales}_{\text{REO}}}{\text{Sales}_{\text{Existing}}}$.3147958	.1648681	.0834633	.795764	97
% Foreclosed	.0870826	.0719943	.0104154	.4205121	97
$\Delta \log (\text{Price})_{03-06}$.2974835	.179294	.0389295	.7288995	97
Share LTV > 80%	.1452959	.0756078	.025514	.3282766	97
Frac Second Mort, 06	.2026752	.0527425	.0259415	.2896224	97
Saiz Land Unav	.2779021	.2112399	.009317	.7964462	97
Wharton Land Reg	.2215807	.7050566	-1.239207	1.89206	97

Note: Summary statistics for variables used in regression analysis. All data is from CoreLogic and fully described in Appendix B.4. Data is for 100 largest MSAs excluding three for which complete data are unavailable as described in the appendix.

downturn more formally, we estimate regressions of the form:

$$\begin{aligned}
Y = & \beta_0 + \beta_1 \max \Delta_{03-06} \log (P) + \beta_2 [\Delta_{03-06} \log (P)]^2 \\
& + \beta_3 (Z \max \text{Share LTV} > 80\%) + \beta_4 (\Delta_{03-06} \log (P) \times Z \text{LTV} > 80\%) \\
& + \beta_5 (Z \% \text{ Second Mortgage, 2006}) + \beta_6 (\Delta_{03-06} \log (P) \times Z \% \text{ Second}) \\
& + \beta_7 (Z \text{ Saiz Land Unavailability}) + \beta_8 (Z \text{ Wharton Land Use Regulation}) + \varepsilon
\end{aligned} \tag{2.2.1}$$

where Z represents a z-score and the outcome variable Y is either the maximum change in log price, the maximum change in log retail prices, the maximum change in log existing home sales, the maximum change in log retail sales, the maximum REO share, or the fraction of houses that experience a foreclosure. The key coefficient is β_4 . This regression is similar in spirit to Lamont and Stein (1999), who show that prices are more sensitive to income shocks in cities with a larger share of high LTV households, except rather than using income shocks to measure volatility, we use the size of the preceding bubble as measured by 2003-2006 price growth. We add the fraction of individuals with a second mortgage or home equity loan to the regression because these loans have received attention in analyses of the downturn (Mian and Sufi, 2011). Finally, to proxy for the housing supply elasticity we use a land unavailability index and the Wharton land use regulation index both from Saiz (2010). Table 2.2.1 shows summary statistics for our left hand side variables in the top panel and our right hand side variables in the bottom panel

The regression results are shown in Table 2.2.2. The first two columns show the impacts on price and retail price. While the additional variables do not explain all of the non-log-linearity, they have substantial predictive power. The key coefficient shows that the interaction between a large bubble and the share of

Table 2.2.2: Cross MSA Regressions on the Impact of the Size of the Bubble and Its Interaction With High LTV

Dependent Variable:	$\Delta \log(P)$	$\Delta \log(P_{\text{Retail}})$	$\Delta \log(\text{Sales}_{\text{Existing}})$	$\Delta \log(\text{Sales}_{\text{Retail}})$	$\frac{\text{Sales}_{\text{REO}}}{\text{Sales}_{\text{Existing}}}$	% Foreclosed
$\Delta \log(\text{Price})_{03-06}$	1.501 (0.656)**	0.884 (0.444)**	-1.932 (0.450)***	-0.493 (0.797)	-1.662 (0.477)***	-0.615 (0.189)***
$\Delta \log(\text{Price})_{03-06}^2$	-3.369 (0.785)***	-2.440 (0.529)***	1.431 (0.573)**	-1.046 (0.936)	3.016 (0.561)***	1.267 (0.229)***
Z Share LTV > 80%	0.062 (0.040)	0.064 (0.030)**	-0.008 (0.038)	0.005 (0.054)	-0.009 (0.033)	-0.008 (0.015)
$\Delta \log(P) \times Z \text{ LTV} > 80\%$	-0.314 (0.155)**	-0.314 (0.135)**	-0.146 (0.138)	-0.379 (0.182)**	0.218 (0.107)**	0.197 (0.067)***
Z Frac Second Mort, 06	-0.058 (0.033)*	-0.037 (0.023)	-0.066 (0.033)**	-0.071 (0.044)	-0.022 (0.027)	-0.019 (0.010)*
$\Delta \log(P) \times Z \text{ Second}$	0.099 (0.112)	0.036 (0.087)	0.102 (0.096)	0.049 (0.121)	0.245 (0.088)***	0.080 (0.036)**
Z Saiz Land Unav	-0.021 (0.018)	-0.008 (0.015)	-0.011 (0.018)	-0.014 (0.021)	0.003 (0.015)	-0.003 (0.006)
Z Wharton Land Reg	-0.031 (0.014)**	-0.027 (0.011)**	0.017 (0.017)	-0.001 (0.019)	0.019 (0.012)	0.008 (0.005)*
Constant	-0.417 (0.107)***	-0.282 (0.073)***	-0.551 (0.064)***	-0.944 (0.132)***	0.461 (0.081)***	0.136 (0.031)***
r^2	0.644	0.703	0.298	0.353	0.521	0.606
N	97	97	97	97	97	97

Note: * = 10% Significance, ** = 5% Significance *** = 1% significance. All standard errors are robust to heteroskedasticity. All data is from CoreLogic and fully described in Appendix B.4. Data is for 100 largest MSAs excluding 3 for which complete data are unavailable as described in appendix B.4. The key take-away is the coefficient on the interaction between bubble size and LTV, which shows that these two factors together were associated with a large price decline in the bust.

homes with high LTV is correlated with large price declines, as suggested by Figure 2.2.2. These interactions also have a large effect on the REO share of sales and fraction foreclosed, suggesting that foreclosures have something to do with these trends. The coefficient on land regulation is also negative yet small, reflecting the amplification provided by a high housing supply elasticity. Having many houses with a second mortgage also reduces prices.

For sales, the regression has noticeably less predictive power and the dominant term is the constant. As discussed in the analysis of the national calibration above, this suggests that foreclosures combined with the size of the bubble will do a much worse job explaining the volume decline than the price decline, something that will be borne out in our cross-MSA simulations. The interaction between LTV and the bubble is insignificant for existing sales but significant and negative for retail sales. The pattern of REO volume largely replacing retail volume is consistent with the four markets with high levels of foreclosure in Figure 2.1.1.

2.3 Housing Market Model

To theoretically examine the effect of foreclosures, we develop a model of the housing market in which foreclosures are exogenous. We subsequently embed this model in a framework in which default is modeled more realistically. Consequently, in this section, we focus on the mechanisms and qualitative predictions and defer a quantitative analysis of the model to section 2.7.

2.3.1 Setup

We consider a Diamond-Mortensen-Pissarides-style general equilibrium search model of the housing market. Search frictions play an important role in housing markets: houses are illiquid, most households own one house and move infrequently, buyers and sellers are largely atomistic, and search is costly, time consuming, and random. Additionally, the outside options of market participants are crucial in search models, so a search framework is well-suited to formalizing the choosey buyer effect described in the introduction.

Time is discrete and the discount factor is β . There are a unit mass of individuals and a unit mass of houses, both fixed. This is a good approximation of the the downturn, in which there has been a very low level of new construction and decreased migration.⁷

The setup of the model's steady state is illustrated schematically in Figure 2.3.1. Table 2.3.1 defines the model's key variables. To simplify the analysis, we assume no default in steady state, which is approxi-

⁷We do not consider the impact of long-run changes in the homeownership rate and retirement rate on the long-run equilibrium of the market, nor do we consider the long-run impact of new construction, both of which may be affected by the downturn and are important subjects for future research.

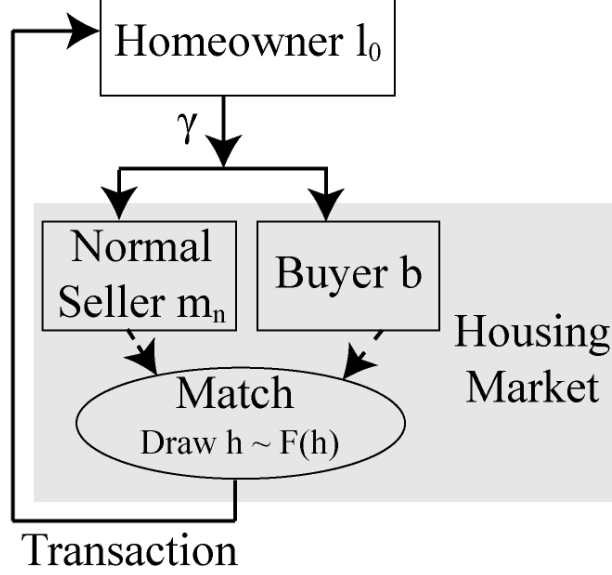


Figure 2.3.1: Schematic Diagram of No Foreclosure Steady State of Model

mately the case when prices are stable or rising.⁸ In steady state, mass l_0 of individuals are homeowners. Homeowners randomly experience shocks with probability γ that induce them to leave their house as in Krainer (2001) and Ngai and Tenreyro (2010).⁹ We assume that these shocks occur at a constant rate and that only individuals who receive a moving shock search for houses. This assumption turns off the amplification channel identified by Novy-Marx (2009) through which endogenous entry and exit decisions by market participants in a search model create a feedback loop which magnifies the effects of fundamental shocks.

An individual who receives a moving shock enters the housing market as both a buyer with flow utility u_b and a normal seller with holding cost m_n . Because shocks create both a buyer and a seller, the model is a closed system with a fixed population.¹⁰ In Section 2.7 we compare our model's predictions to data from both national and local markets, although the model as literally interpreted applies best to an metropolitan area.

As in Ngai and Tenreyro (2010), we assume that the buyer and seller act as independent agents. This means that there is no interaction between the buyer's problem or bargaining game and the seller's, and there is no structure placed on whether an individual buys or sells first. This assumption is not innocuous,

⁸Allowing default in steady state complicates the analysis but does not substantially change the results.

⁹Moving shocks are a reduced form for a number of different life events that trigger a change in housing preference, such as the birth of children, death, job changes, and liquidity shocks.

¹⁰We use a closed system so that housing prices are not determined principally by the flow rates of buyers into and sellers out of the market but rather by the incentives of buyers and sellers in the market. Most moves are within-MSA (Sainai and Souleles, 2009) or to MSAs with highly correlated housing prices, so the assumption of a closed system is reasonable.

Table 2.3.1: Variables in Housing Market Model

Variable	Description
Endogenous Variables	
h	Stochastic Match Quality $\sim F(h)$
h_n, h_d	Cutoff h for normal, REO sellers
$S_{m,h}^B, S_{m,h}^S$	Surplus of type m seller with match quality h for buyer, seller
$p_{m,h}$	Price for type m seller with match quality h
μ	Market tightness (buyers/sellers)
$q_s(\mu), q_b(\mu)$	Prob seller meets buyer, buyer meets seller
r_m, r_d	Ratio of normal, REO sellers to total sellers
l_0, l_1	Masses of homeowners, homeowners that could foreclose
v_b, v_n, v_d, v_r	Masses of buyers, normal sellers, REO sellers, renters
Value Functions	
V_h	Value of owning home with match quality h
V_n, V_d	Value of seller for normal, REO sellers
B	Value of buyer
R	Value of renter
Parameters	
β	Discount factor
γ	Probability of moving shock
α	Probability moving shock causes foreclosure
σ	Probability of leaving renting
θ	Seller's Nash bargaining weight
χ	Probability of match in period (C-D matching function)
ξ	Exponent in C-D matching function
λ	Parameter of exponential distribution for $F(h)$
a	Shifter on exponential distribution for $F(h)$
u_b, u_r	Flow utility of being a buyer, renter
m_n, m_d	Flow utility of being a seller for normal, REO

as whether homeowners have sufficient liquidity to buy first may be important for market equilibrium and may affect bargaining as individuals who buy before selling holding two homes (Anenberg and Bayer, 2012). However, these effects are likely to be small relative to REOs. For instance, Springer (1996) examines several measures of seller motivation and finds that only REO sellers are distinguishable from normal sellers, and Anenberg and Bayer find that individuals who buy first sell their homes at a two percent discount, a figure that is swamped by the average REO discount.

Buyers and sellers in the housing market are matched each period. Matching is entirely random and search intensity is fixed, allowing us to focus on the effects of distressed sales rather than the search mechanism. When matched, the buyer draws a flow utility h from a distribution $F(h)$. Utility is linear and house valuations are purely idiosyncratic so that the transaction decision leads to a cutoff rule. These valuations are completely public, and prices are determined by generalized Nash bargaining. Because buyers know whether the seller is an individual or a bank in practice, symmetric information is reasonable. If the buyer and seller decide to transact, the seller leaves the market and the buyer becomes a homeowner in l_0 deriving

flow utility h from the home until they receive a moving shock. If not, the buyer and seller each return to the market to be matched next period. Note that for simplicity we do not allow speculators or “flippers,” who would presumably sell quickly.

We introduce foreclosures into this basic steady state setup by adding two key ingredients. First, REO sellers have a higher holding costs, which is the case for several reasons. Mortgage servicers, who execute the foreclosure and REO sale, have substantial balance sheet concerns. In most cases, they must make payments to security holders until a foreclosure liquidates, and they must also assume the costs of pursuing the foreclosure, securing, renovating, and maintaining the house, and selling the property (Theologides, 2010). Furthermore, even though they are paid additional fees to compensate for the costs of foreclosure and are repaid when the foreclosed property sells, the servicer’s effective return is likely far lower than its opportunity cost of capital. Additionally, owner-occupants have much lower costs of maintenance and security. Finally, REO sellers usually leave a property vacant and thus forgo rental income or flow utility from the property.

An implicit assumption is that no deep-pocketed and patient market maker buys from distressed sellers and holds the property until a suitable buyer is found. While investors or “flippers” have bought some foreclosures, most have been sold by realtors to homeowners. This is likely due to agency problems and high transactions costs.

Second, individuals who experience a foreclosure are locked out of the market. This reflects the fact that a foreclosure dramatically reduces a borrower’s credit score. Indeed, many banks, the GSEs, and the FHA will not lend to someone who recently defaulted. Instead, foreclosed individuals become renters. This is supported by the data: Malloy and Shan (2011) use credit report data to show that households that experience a foreclosure start are 55-65 percentage points less likely to have a mortgage two years after a foreclosure start. For simplicity, we assume that the rental market is segmented, and renters flow back into buying at an exogenous and fixed rate. While segmentation is a somewhat extreme assumption in the long run, it is a more reasonable approximation for the short-run effects in which we are primarily interested as conversions from owner occupied to rental units are costly and slow. Introducing an endogenous rental price and making the outflow rate covary with the price would create a force that mitigates some of the effects in the model.

Because flow utilities for foreclosures are drawn from the same distribution as for non-foreclosures, we are implicitly assuming that foreclosures are of roughly equal quality, which is likely not the case in practice (Gerardi et al., 2012). In our calibration, we are careful to use moments from studies that carefully control for quality.

While these are the only two new assumptions we make, foreclosures may have other effects. They

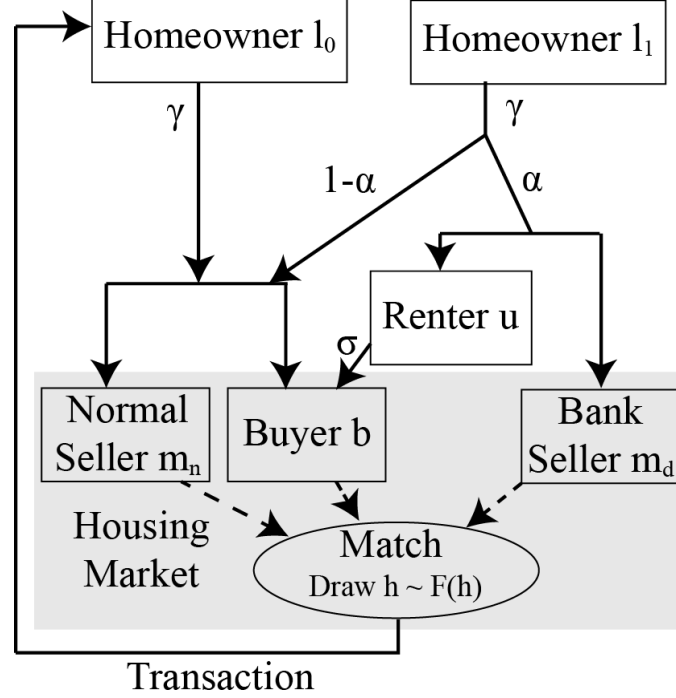


Figure 2.3.2: Schematic Diagram of Housing Market Model With Foreclosure

may cause negative externalities on neighboring properties due to physical damage, the presence of a vacant home, or crime. Campbell et al. (2011) show that such effects are small and highly localized, although contagion is certainly possible in neighborhoods with high densities of foreclosures. There may also be buyer heterogeneity with respect to their willingness to purchase a foreclosure, generating an additional channel through which the REO discount widens as non-natural buyers purchase foreclosures. Finally, foreclosures may cause banks to limit credit supply, as shown theoretically by Chatterjee and Eyigungor (2011).

The two critical assumptions are introduced into the model in Figure 2.3.2. To simplify the analysis, we assume away re-default.¹¹ Instead, we consider a mass of potential defaulters and analyze how these potential defaulters flow through the system. One can think of these potential defaulters as homeowners with high mortgage balances as will be the case in section 2.5. These individuals have mass l_1 , and at time $t = 0$, when we introduce the exogenous foreclosure shock, we move everyone in l_0 to l_1 . Potential defaulters in l_1 also receive moving shocks with probability γ , but if they receive a moving shock it triggers a foreclosure with probability $\alpha(t)$ and is a normal moving shock with probability $1 - \alpha(t)$. If it is a normal moving shock, the homeowners becomes a buyer and a seller as in steady state. A foreclosure shock, however, causes

¹¹We assume away re-default to keep the model consistent with the extended model in section 2.5. While this assumption does slightly increase the speed of convergence back to steady state over the course of the crisis, it does not substantively alter the quantitative or qualitative results.

a bank or GSE with holding cost m_d to take possession of the house and enter the housing market and the homeowner to become a renter with flow utility u_r .¹² Renters become buyers each period with exogenous probability σ . Because there is no re-default, all buyers, including those who were formerly renters, are added to l_0 when they buy a house, so the model gradually returns to steady state.

Buyers and sellers of both types are matched in the housing market. Let $v_b(t)$, $v_r(t)$, $v_n(t)$, and $v_d(t)$ be the masses of buyers, renters, normal sellers, and REO sellers in the market at time t . Market tightness $\mu(t)$ is equal to the ratio of buyers to sellers:

$$\mu(t) = \frac{v_b(t)}{v_n(t) + v_d(t)}. \quad (2.3.1)$$

Unlike general equilibrium search models of the labor market in which market tightness is determined principally by a free entry condition for firms posting vacancies, here market tightness is determined by flows into renting due to default and out of renting at rate σ .

For the matching technology, we use a standard Cobb-Douglas matching function so that the number of matches when there are b buyers and s sellers is $\chi b^\xi s^{1-\xi}$. The probability a seller meets a buyer in a period with market tightness μ is given by $q_s(\mu) = \frac{\chi b^\xi s^{1-\xi}}{s} = \chi \mu^\xi$, and the probability a buyer meets a seller is $q_b(\mu) = \frac{\chi b^\xi s^{1-\xi}}{b} = \chi \mu^{\xi-1}$.

Let $V_h(t)$ be the value of being in a house with match quality h at time t , $V_m(t)$ be the value of being a seller of type m (either n or d) at time t , $B(t)$ be the value of being a buyer at time t , and $R(t)$ be the value of being a renter at time t . $V_h(t)$ is equal to the flow payoff plus the discounted expected continuation value:

$$V_h(t) = h + \beta \{ \gamma [V_n(t+1) + B(t+1)] + (1 - \gamma) V_h(t+1) \}. \quad (2.3.2)$$

The match surplus created when a buyer meets a seller of type $m = \{n, d\}$ and draws an idiosyncratic match quality of h at time t is a key value in the model. Denote this surplus by $S_{m,h}(t)$, the buyer's portion of the surplus by $S_{m,h}^B(t)$, and the seller's portion by $S_{m,h}^S(t)$. Let the price of the house sold if a transaction occurs be $p_{m,h}(t)$. The buyer's share of the surplus is equal to the value of being in the house minus the price and their outside option of staying in the market:

$$S_{m,h}^B(t) = V_h(t) - p_{m,h}(t) - u_b - \beta B(t+1). \quad (2.3.3)$$

¹²The bank must hold a foreclosure auction, but in the vast majority of cases the auction reserve is not met and the bank takes the house as an REO. For instance, Campbell et al. (2011) report that 82 percent of foreclosures in Boston are sold as REOs rather than at auction. For simplicity we assume all houses become REO.

The seller's share of the surplus is equal to the price minus their outside option of staying in the market:

$$S_{m,h}^S(t) = p_{m,h}(t) - m - \beta V_m(t+1). \quad (2.3.4)$$

Prices are set by generalized Nash bargaining with weight θ for the seller, so:

$$\frac{S_{m,h}^S(t)}{S_{m,h}^B(t)} = \frac{\theta}{1-\theta} \quad \forall m. \quad (2.3.5)$$

Buyers and type m sellers will transact if the idiosyncratic match quality h is above a threshold value, corresponding to zero total surplus and denoted by $h_m(t)$. Because total surplus is:

$$S_{m,h}(t) = V_h(t) - (m + u_b) - (\beta B(t+1) + \beta V_m(t+1)) \quad (2.3.6)$$

the cutoff is implicitly defined by:

$$V_{h_m}(t) = m + u_b + \beta (B(t+1) + V_m(t+1)). \quad (2.3.7)$$

We can then define the remaining value functions. The value of being a type m seller is equal to the flow payoff plus the discounted continuation value plus the expected surplus of a transaction times the probability a transaction occurs. Because sellers meet buyers with probability $q_s(\mu(t))$ and transactions occur with probability $1 - F(h_m(t))$, V_m is defined by:

$$V_m(t) = m + \beta V_m(t+1) + q_s(\mu(t)) (1 - F(h_m(t))) E[S_{m,h}^S(t) | h \geq h_m(t)]. \quad (2.3.8)$$

The most important aspect of V_m is that in a downturn $q_s(\mu)$ falls below its steady state value because foreclosures create renters rather than buyers ($\mu < 1$). The chance that a seller does not meet a buyer thus reduces the value of being a seller.

The value of being a buyer is defined similarly, although we must account for the fact that the buyer can be matched with two types of sellers. Let the probability of matching with a type m seller conditional on a match be $r_m(t) = \frac{v_m(t)}{v_n(t) + v_d(t)}$. B is defined by:

$$B(t) = u_b + \beta B(t+1) + q_b(\mu(t)) \sum_m r_m(t) (1 - F(h_m(t))) E[S_{m,h}^B(t) | h \geq h_m(t)]. \quad (2.3.9)$$

Because of random matching, as more REO sellers enter the market the weight on REO sellers in the buyer's value function r_d rises. REO sellers are more likely to sell, so foreclosures raise the value of being a buyer.

The decline in μ caused by foreclosures also raises $q_b(\mu)$, further increasing the value of being a buyer.

It is worth discussing what the implications of allowing buyers to direct their search towards foreclosures would be. A model with completely segmented REO and retail markets produces unreasonable parameter values. Intuitively, the REO and retail markets are linked by a buyer indifference condition that the probability of a match times the surplus must be the same in the REO market and the retail market. With a reasonable foreclosure discount, buyer indifference can only hold if the opportunity cost of waiting slightly longer for a distressed sale – the flow utility from being in that house – is implausibly high.

Furthermore, it is unlikely that any buyers look exclusively at one type of property. Instead, partially-directed search, in which buyers are able to direct their search to particular sub-markets in which the REO share of vacancies is higher than other sub-markets but is still not close to one, is most plausible. Examples of sub-markets include neighborhoods within a MSA or lower priced homes where there are likely to be more foreclosures. In this case, the effects we identify would be most pronounced in those sub-markets which had the highest REO share of vacancies, although there would be some spillovers because marginal individuals would switch to the REO-laden market. This is consistent with the findings of Landvoigt et al. (2012) that price declines in San Diego were stronger at the lower end of the market. We leave understanding the role of foreclosures for within-housing-market dynamics to future research.

The value of being a renter is defined as:

$$R(t) = u_r + \beta \{ \sigma B(t+1) + (1 - \sigma) R(t+1) \}. \quad (2.3.10)$$

We will assume $u_r = u_b$, so that a renter is simply a buyer without the option to buy.

The conditional expectation of the surplus given that a transaction occurs appears repeatedly in the value functions. This quantity can be simplified as in Ngai and Tenreyro (2010) by using (2.3.2) together with (2.3.6):

$$S_{m,h}(t) = V_h(t) - V_{h_m}(t) = \frac{h - h_m(t)}{1 - \beta(1 - \gamma)}. \quad (2.3.11)$$

The conditional expectation is

$$E[S_{m,h}(t) | h \geq h_m(t)] = \frac{E[h - h_m(t) | h \geq h_m(t)]}{1 - \beta(1 - \gamma)}. \quad (2.3.12)$$

We parameterize $F(\cdot) \sim \exp(\lambda) + a$, an exponential distribution with parameter λ shifted over by a constant a . The memoryless property of the exponential distribution implies that $E[S_{m,h}(t) | h \geq h_m^*(t)] = \frac{1}{\lambda}$. This is a fairly strong assumption. By using the exponential distribution in our simulations, we eliminate changes in the expected surplus due to changes in tail conditional expectations of the F distribution, which

cannot be observed.

The model is completed with the laws of motion for the mass of sellers of type m , buyers, renters, and homeowners of type l_i . These laws of motion, which formalize Figure 2.3.2, are in appendix B.1.2.

Prices can be backed out by using Nash bargaining along with the definitions of the surpluses and (2.3.12) to get:

$$p_{m,h}(t) = \frac{\theta(h - h_m(t))}{1 - \beta(1 - \gamma)} + m + \beta V_m(t + 1) \quad (2.3.13)$$

This pricing equation is intuitive. The first term contains $h - h_m(t)$, which is a sufficient statistic for the surplus generated by the match as shown by Shimer and Werning (2007). As θ increases, more of the total surplus is appropriated to the seller in the form of a higher price. This must be normalized by $1 - \beta(1 - \gamma)$, the effective discount rate of a homeowner. The final two terms represent the value of being a seller next period, which is the seller's outside option. These terms form the minimum price at which a sale can occur, so that all heterogeneity in prices comes from the distribution of h above the cutoff $h_m(t)$. Because with the exponential distribution $E[h - h_m(t)] = \frac{1}{\lambda}$, all movements in average prices work through $V_m(t + 1)$.

2.3.2 Numerical Methods

For reasonable parameter values, the model has a unique steady state that can be solved block recursively and studied analytically. The full derivation and existence and uniqueness proofs for the steady state can be found in appendix B.1.1. Although there are no foreclosures in steady state, the price and probability of sale for a REO seller are well defined and represent what would occur if a measure zero mass of normal sellers were instead REO sellers. For a fixed idiosyncratic valuation h , REO properties sell faster and at a discount due to the higher holding costs of distressed sellers.

The dynamics of the model, however, have no analytic solution, so we turn to numerical simulations. We solve the model using Newton's method as described in appendix B.1.2.

Simulating the model requires choosing parameters. We defer a more rigorous quantitative analysis to section 2.7, which features a richer model, and focus on the mechanisms at work in this section. Consequently, for now we present simulation results using an illustrative calibration similar to the one described in section 2.5. We simulate a wave of foreclosures by moving everyone in l_0 to l_1 at time $t = 0$ and raising α for a period of five years. After the wave of foreclosures, the model returns to the original steady state.

2.4 Basic Model Results and Mechanisms

2.4.1 Market Tightness, Choosey Buyer, and Compositional Effects

The qualitative results in our model are caused by the interaction of three different effects: the “market tightness effect,” the “choosey buyer effect,” and the “compositional effect.” Each is crucial to understand the effect of foreclosures on the housing market.

First, because foreclosed individuals are locked out of the housing market as renters and only gradually flow back into being buyers, foreclosures reduce market tightness $\mu(t)$. This mechanically decreases the probability a seller meets a buyer in a given period and triggers endogenous responses as each party’s outside options to the transaction changes, altering the bargaining and the h s for which a sale occurs. For sellers, the reduction in market tightness reduces the value of being a seller for both types of seller, reducing prices and causing sellers to sell more frequently. The endogenous response is stronger for REO sellers who have a higher opportunity cost of not meeting a buyer. For buyers, the elevated probability of meeting a seller raises their expected value, leading to lower prices and a shift in the cutoffs that makes buyers more choosey.¹³ The market tightness effect elucidates that an important element of foreclosures is a reduction in demand relative to supply, as in a typical market a move creates a buyer and a seller while foreclosures create an immediate bank seller but a buyer only when the foreclosed upon individual’s credit improve. This contrasts with some market analysts who treat foreclosures as a shift out in supply rather than a reduction in today’s demand.

Second, the value of being a buyer rises because the buyer’s outside option to transacting, which is walking away and resampling from the distribution of sellers next period, is improved by the prospect of finding an REO seller who will give a particularly good deal. Mathematically, as REOs make up a larger fraction of total vacancies, r_d rises and the term in the sum in (2.3.9) relating to REO sales gets a larger weight. This term is larger because REO sellers are more likely to transact both in and out of steady state. The resulting increase in buyers’ outside options leads buyers to become more aggressive and demand a lower price from sellers in order to be willing to transact. In equilibrium, this leads to buyers walking away from more sales. Importantly, this effect will be most prevalent in the retail market where sellers are less desperate and therefore less willing to accommodate buyers’ demand for lower prices, resulting in a freezing up of the retail market.

The choosey buyer effect is new to the literature. Albrecht et al. (2007, 2010) introduce motivated sellers into a search model, but focus on steady-state matching patterns (eg whether a high type buyer can match

¹³In the calibration utilized here and in later sections, we set $\xi = .84$. Thus the effect on $q_s(\mu)$ significantly outweighs the effect of market tightness on $q_b(\mu)$.

with a low type seller) and asymmetric information regarding seller type. Duffie et al. (2007) consider a liquidity shock similar to our foreclosure shock, but a transaction occurs whenever an illiquid owner meets a liquid buyer, and so while there are market tightness effects their model does not have a choosey buyer effect.

The market tightness effect and choosey buyer effect are mutually reinforcing. As discussed above, the market tightness effect is more pronounced for REO sellers. Because the value of being an REO seller falls by more, REO sellers become even more likely to sell relative to non-REO sellers during the downturn. This sweetens the prospect of being matched with an REO seller next period, amplifying the choosey buyer effect.

Finally, a greater share of REO sales makes the average sale look more like REO properties, which sell faster and at lower prices both in and out of steady state. Foreclosures thus cause a mechanical compositional effect that affects sales-weighted averages such as total sales and the aggregate price index.

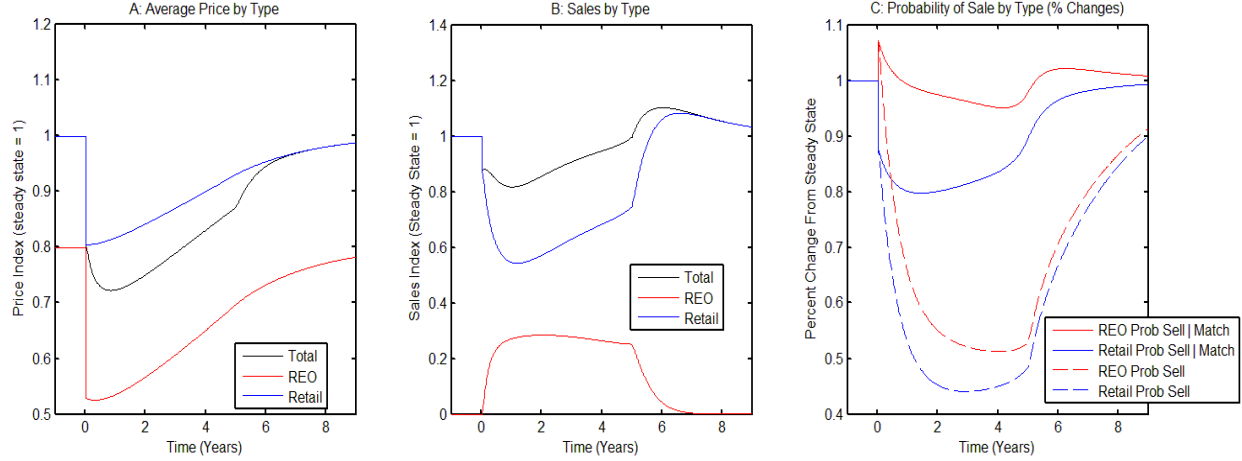
The market tightness effect is the aspect of the model that comes closest to a standard Walrasian analysis with a single market for housing. By reducing the number of buyers relative to sellers, it is similar to an inward shift in the demand curve relative to the supply curve that reduces both prices and transaction volume. The market tightness effect does, however, asymmetrically impact REO and retail sellers due to their differential holding costs, leading to a greater freezing up of the retail market as buyers walk away from retail sellers in hopes of contacting increasingly-desperate REO sellers. These types of differential effects and further feedback loops – which stem from the choosey buyer effect and its interaction with the market tightness effect – are novel to the literature and differentiate our model from a simpler Walrasian model.

Furthermore, all three effects dissipate more slowly than in traditional asset pricing models because they depend on flows as well as stocks and lead to a sluggish return of the housing market to steady state. The choosey buyer and compositional effects last as long as foreclosures remain in the market, which is only a few months after the shock ends as these houses sell quickly. However, the market tightness effect persists for much longer as it takes several years for the renters to return to being homeowners.

2.4.2 Qualitative Results

Figure 2.4.1 shows the effect of a the five-year wave of foreclosures. Because the model is entirely forward looking, prices and probability of sale conditional on a match fall discretely on the impact of the shock at $t = 0$. This is typical in completely forward-looking models. The sluggish adjustment of house prices to shocks remains a puzzle for much of the literature, and a solution to this problem is outside the scope of this paper.

As shown in panel A, at $t = 0$ prices fall considerably for both REO and retail and gradually return to



Note: This figure shows the results of the housing market model with exogenous foreclosures using an illustrative calibration similar to the one developed in Section 2.5 and a five-year foreclosure shock. Panels A and B show the average price and sales by type, with pre-downturn price and volume normalized to 1. Prices drop discretely at time zero as is standard in forward-looking models with no uncertainty. The REO discount widens, the aggregate price index is pulled towards the REO index as REOs make up a greater share of the market, and prices rise in anticipation of the end of the downturn. Retail volume plunges dramatically, but the decline is partially made up for by surging REO volume. Panel C shows the probability of sale conditional on a match and the unconditional probability of sale for each type with the pre-downturn probability normalized to one. This panel illustrates the mechanisms at work in the model, as described in the main text. The key take-away is that the probability of sale conditional on a match, which is the clearest indicator of how the behavioral responses of buyers and sellers play out in equilibrium, falls dramatically for retail and is roughly flat for REO.

Figure 2.4.1: Housing Market Model: Qualitative Results

steady state over the next several years. The overall sales-weighted price index dips more than retail sales as foreclosures are averaged in. The price movements lead to a substantial rise in the average REO discount that falls off over time.

Prices fall due to all three effects. Recall that from (2.3.13), movements in the average price of properties sold by a type m seller are controlled by movements in $V_m(t+1)$. The market tightness effect has a direct effect on the value of being a seller and thus brings down prices. Because this effect is stronger for REO sellers, this contributes to the larger REO discount. The choosy buyer effect has an indirect effect on $V_m(t+1)$, as in general equilibrium increased buyer choosiness reduces the value of being a type m seller, which causes prices to fall. The effect of market tightness on the value of being a buyer operates in a similar manner. Finally, there is a pure compositional effect as REO sales become a greater share of total sales, which is shown graphically by the departure of the aggregate price index from the price index for retail sales.

As for sales, the wave of foreclosures sales causes the retail market to freeze up, with retail volume falling substantially as shown in panel B and REOs constituting a larger fraction of total sales than of total vacancies. Total volume, however, does not fall as much because much of the decline in retail sales is offset by REO sales. After the foreclosures end, sales return back to normal in a matter of months as REOs are eliminated from the market. Most of the sluggish adjustment comes from the dissipation of the accumulated

renters and retail sellers, which takes several years.

The intuition behind the effects on transaction volume is more nuanced as the market tightness, choosy buyer, and composition effects have cross-cutting impacts. Panel C, which shows percent changes from steady state in the probability of sale both raw and conditional on a match, elucidates the role of each effect.¹⁴

Consider first the probability of sale conditional on a match, controlled by $h_m(t)$.¹⁵ The market tightness effect on the probability a seller meets a buyer raises the probability of sale conditional on a match because sellers meet buyers less frequently and thus have a greater incentive to sell when they are matched, an effect which is stronger for REO sellers. The choosy buyer effect and the effect of market tightness on the probability a buyer meets a seller both reduce the probability of sale conditional on a match as buyers become more choosy. Panel C shows that the two effects offset for REO sales as the probability of sale conditional on a match fluctuates around its steady state value, while the choosy buyer effect and the market tightness effect on buyers dominate for retail sales as the probability of sale conditional on a match falls substantially. The relative strength of these two effects for the two types of sellers thus plays an important role in freezing up the retail market.

The market tightness effect, however, plays an additional role: it mechanically reduces volume because there are fewer buyers. This causes the unconditional probability of sale and thus transaction volume to fall for both types, although it falls more for REO sellers. Note, however, that decline for retail sales is quicker and the trough lasts longer.

The compositional effect also plays an important role in determining transaction volume. Because REOs sell faster both in and out of steady state, as the average sale looks more like an REO, volume rises. This is the main reason why total volume does not fall so dramatically. It is possible for volume to rise, although for reasonable calibrations we find that the market tightness effect is strong enough relative to the compositional effect that REO sales do not make up the full shortfall in retail sales and overall volume falls.

Qualitatively, the model explains many salient features of the housing downturn. The substantial decline in both retail and REO prices is consistent with the data in Figure 2.1.1, and the widened distressed sale discount in a downturn is corroborated by Campbell et al. (2011). The freezing up of the retail market and the large share of REO sales in total sales relative to listings is borne out in the data, as are a rise in times to sale and increasing vacancy rates. The fact that REO sales replace a good deal of the lost volume in the retail market is consistent with the evidence from the hardest hit markets as shown in Figure 2.2.1.

¹⁴The probability of sale conditional on a match is $\exp(-\lambda(h_m(t) - a))$ and the total probability of sale is $q_s(\mu(t)) \exp(-\lambda(h_m(t) - a))$

¹⁵Time to sale is inversely related to the unconditional probability of sale.

2.4.3 Isolating the Role of Each Effect

To further illustrate how each effect contributes to our results, Figure 2.4.2 depicts simulations identical to our main results for a wave of foreclosures except with the market tightness effect, choosey buyer effect, and both the choosey buyer and market tightness effects shut down. Although the market tightness effect plays an outsized role, all three effects are necessary for our results.

The market tightness effect generates a significant fraction of the price and volume declines. Row two also illustrates that the market tightness effect increases the conditional probability of sale for REO sellers during the downturn. Market tightness effects also cause total volume to decline because of the mechanical decrease in matching probabilities.

However, the choosey buyer effect plays an essential role in freezing up the retail market. As can be seen from row two, with no choosey buyer effect the conditional probability of sale for retail sellers essentially remains flat. On the other hand, from row one we can see that when only the choosey buyer effect is present there is a non-trivial decrease in this conditional probability. This freezing up is even more pronounced when both market tightness and choosey buyer effects are present due to their interaction.

The compositional effect mainly reduces the aggregate price index, as shown in row 3 of Figure 2.4.2. It also increases total volume slightly because REO sales sell faster.

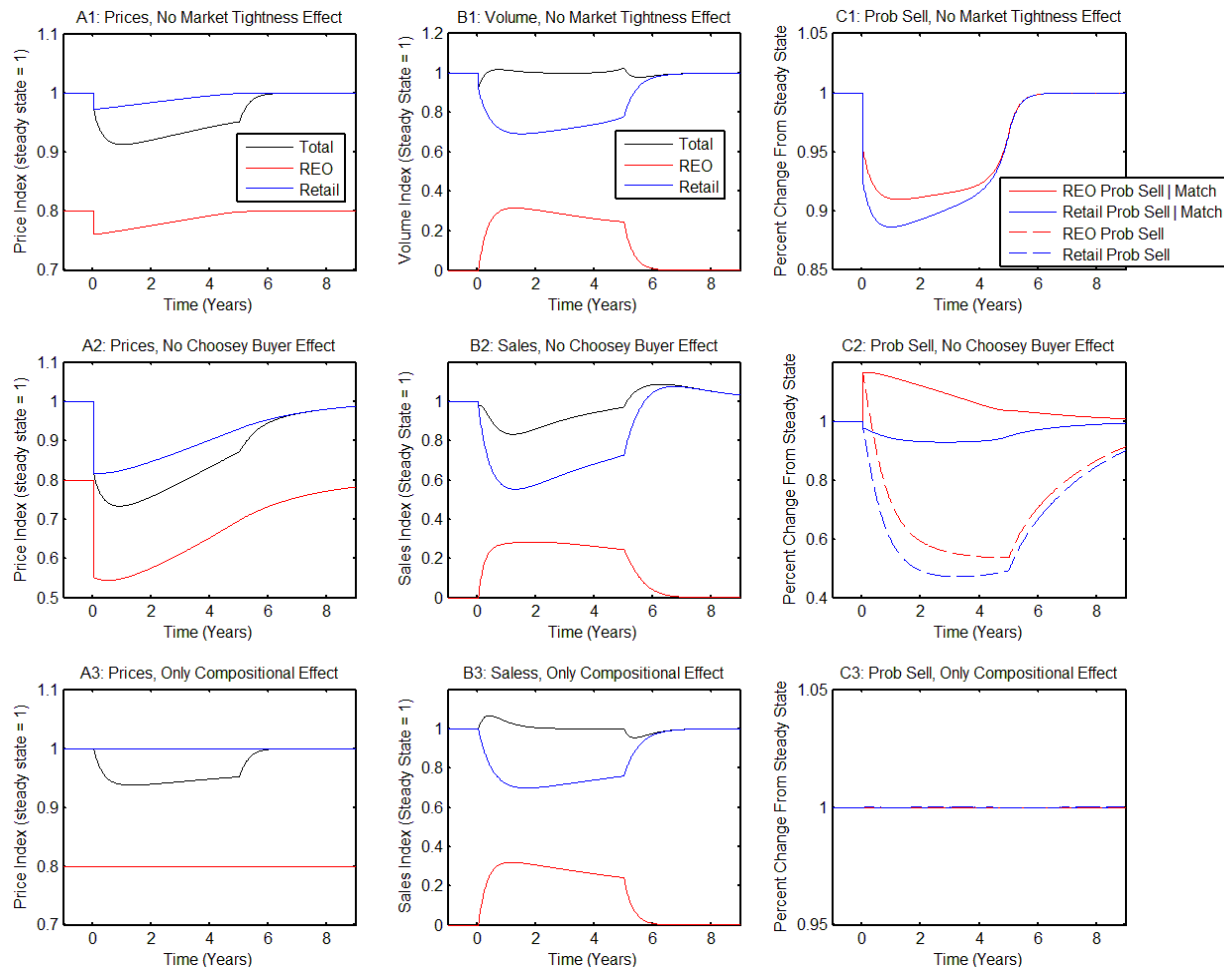
2.5 An Extended Model of Default

Foreclosures are not random events. With few exceptions, negative equity is a necessary but not sufficient condition for foreclosure (Foote et al., 2008). This is because a homeowner with positive equity can sell his or her house, pay off the mortgage balance, and have cash left over without having to default. Homeowners with negative equity, however, are not able to pay the bank and thus default if they experience with a liquidity shock.

The previous section showed that foreclosures have general equilibrium effects that cause prices – and thus homeowner equity – to fall. In a world in which negative equity leads to foreclosure, this will cause more foreclosures and price declines, generating a feedback loop that amplifies the effects of an initial decline in house prices.¹⁶

In this section we embed the housing market component of the exogenous default model developed in the section 2.3 into a model in which negative equity is a necessary but not sufficient condition for default.

¹⁶In all cases we have considered, each additional round of feedback is smaller than the previous one generating a convergent series and a unique dynamic equilibrium, although in principle the feedback could be strong enough to generate a divergent series.



Note: The top row shows price, sales, and conditional and unconditional probability of sale, all normalized to 1 for their pre-downturn values, for the case of no market tightness effect. The second row shows the same results for no choosey buyer effect. The third row shuts down both, leaving only the compositional effect. The figure shows that both the market tightness and outside option effects are critical for the qualitative results; in particular panel C2 shows that without the choosey buyer effect the probability of sale conditional on a match does not fall much for retail sellers. To shut down the market tightness effect, instead of creating renters we instead assume that distressed sale shocks create REO sellers and home-buyers. To shut down the choosey buyer effect, we modify the buyer's value function so that agents behave as if the probability they will hit a distressed seller is zero regardless of the presence of distressed sellers in the market. Because we calibrate to a steady state with no distressed sales, the steady state of these modified models replicates the steady state of our full model. See appendix B.2.1 for full details on these models.

Figure 2.4.2: Isolating the Role of Each Effect

Subsequent sections provide a rigorous quantitative analysis of the extended model and analyze welfare and foreclosure policy using the model.

2.5.1 Default in the Extended Model

We model default as resulting primarily from shocks that cause homeowners with negative equity to be unable to afford their mortgage payments, the so-called “double trigger” model of mortgage default. While “ruthless” or “strategic default” by borrowers has occurred, much of the literature on default argues that strategic default has contributed surprisingly little to foreclosures, particularly at low levels of negative equity.¹⁷ Bhutta et al. (2011) use a method of controlling for income shocks to estimate that the median non-prime borrower does not strategically default until their equity falls to negative 67 percent. Even among non-prime borrowers in Arizona, California, Florida, and Nevada who purchased homes with 100 percent financing at the height of the bubble – 80 percent of whom defaulted within 3 years – over 80 percent of the defaults were caused by income shocks. Similarly, Foote et al. (2008) show that in the Massachusetts housing downturn of the early 1990s, the vast majority of individuals who default have negative equity but most individuals with negative equity do not default. Consequently, the largest estimate of the share of defaults that are strategic is 15 to 20 percent.¹⁸ To keep the model tractable, we thus do not model strategic default, nor do we model the strategic decision of the bank to foreclose or short sales.¹⁹

Modeling negative equity requires that homeowners have loan balances. We assume that homeowners in l_1 have a distribution of loan balances L defined by a CDF $G(L)$.²⁰ So that no foreclosures occur without an additional shock, in general we assume that $G(L)$ has continuous support on $[0, V_n^*]$, where the steady state value of being a normal seller which is equal to the expected price net of the costs of sale. We assume away re-default so that we do not need to worry how new home purchases affect $G(L)$.

To incorporate liquidity shocks into our model, we assume that they occur to individuals with negative equity at Poisson rate γ_I . All other shocks are taste shocks that occur at Poisson rate γ , so that liquidity shocks are in addition to normal shocks.

Liquidity and taste shocks have different effects depending on the equity position of the homeowner.

¹⁷Relevant papers than analyze the default decision and conclude that a “ruthless exercise” option model of default is insufficient include Deng et al. (2000), Bajari et al. (2009), Elul et al. (2010), and Campbell and Cocco (2011).

¹⁸This estimate comes from Experian-Oliver Wyman. Guiso et al. (2009) analyze a survey that asks people whether they strategically defaulted and find that 26 percent of defaults are strategic.

¹⁹“Fishing” – that is listing a home for a high price and hoping that someone who overpays for it will come along as in Stein (1995) – and short sales are unusual because they require sellers to find a buyer who will pay a minimum price, which affects bargaining. Modeling short sales and their effect on market equilibrium is an important topic for future research.

²⁰We are agnostic as to the source of the loan balance distribution and leave this unmodeled. $G(L)$ is fixed over time because principal is paid down slowly, particularly by those in the upper tail of the loan balance distribution who are relevant for the size of the feedback loop.

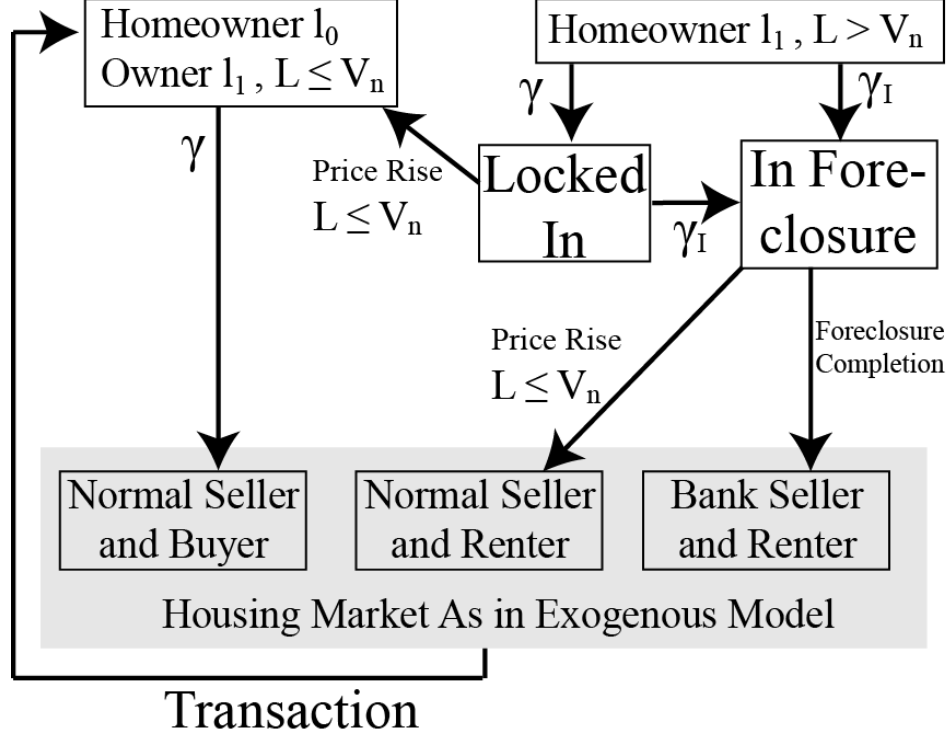


Figure 2.5.1: Extended Model Schematic Diagram

Homeowners with any shock with $L \leq V_n(t)$ have positive equity enter the housing market as a buyer and seller. To keep the model tractable, we assume that buyers and sellers are identical once they pay off their loan balance. Homeowners with $L > V_n(t)$ have negative equity net of moving costs and default if they experience an income shock because they cannot pay their mortgage or sell their house. Defaulters enter the foreclosure process.²¹ Although foreclosure is not immediate and some loans in the foreclosure process do “cure” before they are foreclosed upon, for simplicity we assume that foreclosure occurs immediately. We alter this and introduce foreclosure backlogs in Section 2.8. We also assume that income shocks are a surprise, so an underwater homeowner expecting an income shock cannot list their house with the hope of getting a high-enough price that they can pay off their loan before the bank forecloses. While this may happen infrequently, it is unlikely very unlikely that a desperate seller would receive such a high price.

Finally, homeowners with negative equity who receive a taste shock would like to move but owe more than their house is worth. Consequently, they are “locked in” their current house until prices rise to the point that they have positive equity.²² We assume that once they do not move when they get a taste

²¹It is also possible for banks to possess the house and rent it to the homeowner or to offer a short sale, in which the bank accepts a sale at a price below the outstanding loan balance. While these options have become more popular in recent years due in large part to political pressure, for most of the crisis the banks simply foreclosed on borrowers.

²²Formally, define $w(t)$ as the mass of individuals who are locked in at time t . The distribution of loan balances in $w(t)$ will be the same as $G(L)$ truncated below at $V_n(t)$.

Table 2.5.1: Variables Used In Extended Model

Variable	Description
Endogenous Variables	
L	Loan Balance $\sim G(L)$
w	Mass of locked in homeowners
f	Mass of homeowners in foreclosure process
Parameters	
γ_I	Probability of a liquidity shock
δ	Probability above-water homeowner becomes renter
b_a, b_b	Parameters of Beta distribution for $G(L)$

shock, these homeowners make accommodations and thus do not immediately move when they reach positive equity. Instead, once they have positive equity they become indistinguishable from households in l_0 who are waiting for a moving shock. In our context, lock-in further reduces sales in markets with many foreclosures.

Figure 2.5.1 shows our formalization of default in the extended model in a schematic diagram. The extended model only alters the mechanisms through which homeowners default and enter the housing market, and consequently the housing market component of the model is exactly as in section 2.3 and is thus not depicted in Figure 2.5.1. This structure preserves all of the key intuitions developed in section 2.4. Because there is still no default in steady state, the steady state remains the same. Because the modified model has an identical housing market to the model in section 2.3, the Bellman equations and cutoff conditions are unchanged. The new laws of motion are in Appendix B.1.4, and the additional parameters introduced in the extended model are listed in Table 2.5.1.

2.5.2 Starting the Downturn

An exogenous shock is required to generate an initial price drop. We introduce the exogenous shock in two different ways.

First, we assume that due to both tighter lending practices and income shocks a fraction $\delta(t)$ of individuals who sell their house as a normal seller after receiving a taste shock cannot buy a house and instead transition from owning to renting. This generates an initial market tightness effect that reduces prices, putting some individuals underwater and triggering a price-default spiral. Such a shock fits most naturally in the model. We use a 5-year increase in $\delta(t)$ to perform a sensitivity analysis in section 2.6.

However, when we take the model to the data in section 2.7, it is clear that the main shock in the recent episode is a bursting housing bubble. Our primary interest is understanding the amount of overshooting of prices caused by the presence of foreclosures and not realistically modeling the bursting of the bubble absent foreclosures. Consequently, we introduce a bursting bubble into the model with a permanent decline in a , the minimum flow utility of housing. While the flow utility of housing clearly did not fall overnight, the

source of the initial drop in prices is immaterial to our results and reducing a is the simplest way to generate such a price drop. The reduction in a should thus be seen as a stand-in for a number of factors that have depressed home prices, from changes in credit availability to belief disagreements to irrational exuberance.²³

With a permanent decline in a and a constant hazard of an income shock, all individuals whose loan balance is above the new steady-state price level will eventually default. This does not seem realistic – many homeowners will eventually pay down their mortgage and avoid default. Rather than modeling the dynamic deformation of the $G(\cdot)$ distribution over time, we instead assume that after 5 years the hazard of income shocks γ_I gradually subsides over the course of a year.²⁴

With both exogenous shocks, defaults due to negative equity and the resulting foreclosures amplify the effects of shocks in the housing market due to a price-foreclosure feedback loop. Due to the forward-looking nature of agents in the model, this spiral is capitalized into the prices of retail sales and REO sales when the shock occurs, with further gradual declines as the REO share of volume in the market increases.

2.6 Quantitative Analysis: Calibration and Amplification

Having analyzed the mechanisms at work in our extended model, we now turn to the model's implied magnitudes by conducting a quantitative analysis. In this section, we calibrate the model and examine the strength of the amplification channel. In section 2.7, we take the model to the data on the ongoing downturn.

2.6.1 Calibration

In order to simulate the model numerically, we must choose parameters. As mentioned previously, we parameterize the distribution of idiosyncratic valuations $F(\cdot)$ as an exponential distribution with parameter λ shifted by a . We parameterize the loan balance distribution $G(L)$ as a beta distribution with parameters b_a and b_b scaled to have support on $[0, V_n^*]$. This flexible distribution allows us to approximate the loan balance distribution in various locations on the eve of the crisis as described below. This gives 12 exogenous parameters to calibrate for the basic housing market model – a , γ , m_d , m_n , u_r , u_b , θ , β , σ , λ , χ , and ξ – and three parameters to calibrate unique to the extended model – γ_I , and b_a and b_b . We also must choose the initial shock.

Our calibration procedure proceeds in three steps. We take care to calibrate to pre-downturn moments

²³ While a bursting bubble is the most likely source of a large price change that would put many homeowners underwater, the type of foreclosure crisis we describe could be created by any type of large negative price shock.

²⁴ Formally, after 5 years γ_I falls by 5% of its previous value every month, taking roughly a year to settle at zero.

whenever possible in order to make our tests out-of sample, although in some cases we have no choice but to choose a parameter using data from the housing bust. First, we set γ , u_r , u_b , β , χ , ξ , and γ_I independently to match several moments. Second, we choose a , m_n , m_d , θ and λ so that the steady state of the model matches additional targets. Third, we calibrate b_a and b_b to the appropriate geographic unit that we are considering.

This leaves two variables that we do not choose through calibration: σ , the probability of leaving the rental market, and the initial shock. Although there are several guidelines regarding how long banks deny mortgages to individuals who default,²⁵ there is no good data on this parameter in practice. Consequently, we pursue a two-pronged approach. First, to understand the impact of σ and the exogenous shock, we examine the response of the model to different values of each. Second, we use data on the size of the bust across housing markets to select a preferred calibration of these two parameters, as described in Section 2.7. In the remainder of this section, we describe the moments to which we calibrate the model in our three steps.

Step 1: Exogenous Parameter Choices

We choose u_r , u_b , β , and γ so that one period is equivalent to one week, although the results are insensitive to the time interval chosen. We set the annual discount rate equal to 5%, so that the discount factor is $\beta = 1 - \frac{.05}{52}$. We assume $u_r = u_b = 0$ so that buyers and renters are identical in their flow utility. Buyers, however, have the option to buy which has considerable value so $B > R$. This assumption is equivalent to assuming that buyers and renters pay their flow utility in rent in a perfectly competitive rental market.

We set the probability of moving houses in a given week to fit a median tenure for owner occupants of approximately 9 years from the American Housing Survey from 1997 to 2005, so $\gamma = \frac{.08}{52}$.

We set $\xi = .84$ using estimates from Genesove and Han (2012), who use National Association of Realtors surveys to estimate the contact elasticity for sellers with respect to the buyer-to-seller ratio. χ is then a constant chosen to make sure the probability of matching never goes above 1 for either side of the market. We choose $\chi = .5$, which in our simulations leads to matching probabilities on $[0, 1]$. The results are robust to alternate choices of χ .

The one parameter that we need data from the downturn to choose is γ_I . We set $\gamma_I = 8.6\%$ annually using national data from CoreLogic on the incidence of foreclosure starts for houses with negative equity as described in appendix B.4.

Step 2: Matching the Pre-Downturn Steady State

We then fit the following five aggregate moments from the housing market prior to the housing bust to

²⁵Three years of good credit are needed to get a Federal Housing Administration loan, and according to Fannie Mae guidelines issued in 2010, individuals are excluded from getting a mortgage for two to seven years if they are foreclosed upon, depending on the circumstances. However, these guidelines are not ubiquitous.

the model's steady state to set θ , a , λ , m_n , and m_d :

1. The mean house price for a retail sale, which we set equal \$300,000 as an approximation to Adelino et al.'s (2012) mean house price of \$298,000 for 10 MSAs. In reporting results, we normalize this initial house price to 1. Our results are approximately invariant to the mean house price as long as the residual variance is rescaled proportionally.
2. The residual variance of house prices due to the idiosyncratic preferences of buyers. We set this equal to \$10,000.
3. The REO discount in terms of mean prices, which we set equal to a quality-adjusted 20% based on the average discount in good times from Campbell et al. (2011), whose results are consistent with a literature surveyed by Clauretie and Daneshvary (2009). In Section 2.6 we also consider a 10% discount, closer to the estimates of Zillow (2010) and Clauretie and Daneshvary (2009).
4. Time on the market for retail houses, which we set to 26 weeks as in Piazzesi and Schneider (2009). This number is a bit higher than some papers that use Multiple Listing Service Data such as Anenberg (2012) and Springer (1996), likely because of imperfect adjustment for withdrawn listings and re-listings. Our results are not sensitive to this number.
5. Time on market for REO houses, which we set to 15 weeks. Springer (1996) analyzes various forms of "seller motivation" such as relocation and financial distress using data from Texas from 1991-3. He finds that a foreclosure sales are the only motivated sellers that have significantly reduced time on the market. His estimate is that time on the market is reduced by .2135 log points or 23.7%. However, REO sales are almost never withdrawn from the market, whereas retail sales are frequently withdrawn (Anenberg, 2012). We also attempt to adjust so our number excludes extremely rundown properties that sit on the market for several years.

These moments provide a unique mapping to θ , a , λ , m_n , and m_d , as described in appendix B.1.3.

The calibration procedure results in the parameter values listed in Table 2.6.1, with all prices and dollar amounts in thousands of dollars. Two main things are of note about the calibration. First, $m_d < m_n$, so the flow cost of being a REO seller is higher than the flow cost of being a regular seller. This is due to the fact that distressed sales take less time to sell and trade at a discount in steady state. Second, θ is quite low in order to rationalize the 20 percent discount for REO sales in steady state. This means the buyer will get a majority of the surplus and the value of being a buyer in the buyer's market of the downturn will be high.²⁶

²⁶A low θ is consistent with the logic of directed search and Genesove and Han's (2011) estimate of $\xi = .84$. In directed

Table 2.6.1: Calibrated Parameter Values

a	γ	m_d	m_n	u_r	u_b	θ
3.399	$\frac{0.08}{52}$	-0.403	-0.091	0	0	0.087
β	λ	χ	ξ	γ_I	b_a	b_b
$1 - \frac{0.95}{52}$	3.490	0.5	0.84	$\frac{.086}{52}$	0.898	1.223

Step 3: Geographically-Specific Parameters

To calibrate the two parameters of the loan balance distribution b_a and b_b at the national and MSA level we use proprietary data from CoreLogic on seven quantiles of the combined loan-to-value distribution for active mortgages in 2006. These LTV estimates are compiled by CoreLogic using public records, with the LTV estimates supplemented using CoreLogic’s valuation models.²⁷ Because our model concerns the entire owner-occupied housing stock and not just houses with an active mortgage, we supplement the CoreLogic data with the Census’ estimates of the fraction of owner-occupied houses with a mortgage from the 2005-2007 American Community Surveys. We construct the empirical CDF of the loan balance distribution and then fit a beta distribution with parameters b_a and b_b to the empirical distribution using a minimum distance method described in Appendix B.4. The fit is quite good across the 50th to 100th percentiles of the LTV distribution.] Table 2.6.1 shows the resulting b_a and b_b for the nationwide loan balance distribution.

2.6.2 Strength of Amplification Channel and Comparative Statics

Having calibrated the model, we now gauge the potential magnitude of the amplification channel and elucidate key comparative statics in the extended model. Our initial shock will be one in which a fraction $\delta(t)$ of individuals who receive a taste shock transition from owning a home to renting, as described previously. This causes an initial price decline since it reduces the number of buyers but not the number of sellers. In particular, we use a shock of $\delta(t) = .10$ for five years.²⁸ Holding the initial shock constant, we then vary the average weeks out of the market for a renter, the steady state discount on REO sales, and the loan balance distribution. Specifically, we consider average times out of the market of 1, 1.25, 1.5, and 1.75 years and REO discounts of 10% and 20%. For the loan balance distribution, we consider beta distributions fitted to

search models θ is determined endogenously through price posting behavior. The buyer’s bargaining power is the elasticity of contacts with respect to market tightness ξ and so θ should be below $1 - \xi = .16$. Intuitively the seller has very low bargaining power because the seller’s contact rate is more sensitive to a change in market tightness than the buyer’s.

²⁷Although the CoreLogic estimates of negative equity and the loan-to-value distribution are most cited, some have argued that they do not fully capture the extent of negative equity. Recent estimates by Zillow use credit report data instead of public records and get a higher figure for negative equity than CoreLogic (methodological differences are described here <http://blogs.wsj.com/developments/2012/05/24/negative-equity-more-widespread-than-previously-thought-report-says/>). Beyond issues of data sourcing, Kortewig and Sorensen (2012) argue that traditional methods of estimating house price indices under-estimate the variance of the house price distribution and thus under-estimate the number of loans with high LTV.

²⁸Raising the size of the initial shock within reasonable parameter ranges magnifies the strength of the amplification channel, but the increase is mild.

Table 2.6.2: Sensitivity Analysis: Time Out of Market For Renters

	Price Index Decline	Total Sales Decline
$\sigma = 1/52$ (1 year out)	4.0%	4.4%
$\sigma = 1/65$ (1.25 years out)	4.9%	4.8%
$\sigma = 1/78$ (1.5 years out)	5.8%	5.2%
$\sigma = 1/91$ (1.75 years out)	6.6%	5.5%

match the national data as well as data from New York, which had a low share of high LTV homes, and Las Vegas, which had a high share.

Since our initial shock is a market tightness effect, the strength of this effect will be governed by σ . Table 2.6.2 reports the price index decline and total volume decline generated by the initial shock in the absence of any defaults.

Increasing the average length of time for which individuals who transition to renting stay out of the housing market leads to greater percentage decreases in both the price index and total volume. The effects of such a shock as small as the one we consider are relatively modest, but our key question remains the potential strength of the amplification channel when we allow for defaults. Table 2.6.3 reports the results from varying the REO discount and loan balance distribution in addition to σ . Rather than reporting levels, each entry reports the percentage amplification of the aggregate price index decline and sales decline generated by defaults over and above the decline created by the initial shock in a price-volume pair.

Table 2.6.3 shows that foreclosure spirals can significantly amplify an initial shock. The strength of this spiral grows as we increase the amount of time individuals who default spend out of the housing market due to a more persistent market tightness effect. The shape of the loan balance distribution also plays a critical role in determining the strength of the amplification. A greater proportion of individuals with high LTV ratios implies that a given initial shock will put a greater fraction of the market underwater. This leads to greater numbers of foreclosures, more powerful general equilibrium effects, and in turn even more foreclosures. This illustrates the fragility created by the combination of a housing bubble and reduced down payment requirements.

Table 2.6.3 also shows that a lower steady state REO discount dampens the spiral. This is in part due to a compositional effect which ameliorates the effect of a given number of REOs on the price index. Additionally, though, the choosey buyer effect is weaker since waiting for an REO sale is no longer as attractive, and so the retail market does not freeze up as much. Finally, the amplification channel can generate significant *total* volume declines, greater than we saw in section 2.3 with exogenous defaults for a given shock size. The reason is that in the extended model, relative to section 2.3, the number of individuals who become locked-in during the downturn can be substantial. Note that this implies a greater REO share of vacancies which

Table 2.6.3: Sensitivity Analysis: Loan Balance Distribution and REO Discount

$\sigma = 1/52$	REO Disc. 10%	REO Disc. 20%	$\sigma = 1/65$	REO Disc. 10%	REO Disc. 20%
New York	(4.8%, 2.8%)	(6.2%, 3.4%)	New York	(6.6%, 4.0%)	(8.4%, 4.9%)
National	(22.3%, 11.9%)	(29.4%, 14.6%)	National	(30.0%, 17.3%)	(38.5%, 20.5%)
Las Vegas	(46.0%, 23.7%)	(60.2%, 28.7%)	Las Vegas	(62.0%, 34.6%)	(79.3%, 40.6%)
$\sigma = 1/78$	REO Disc. 10%	REO Disc. 20%	$\sigma = 1/91$	REO Disc. 10%	REO Disc. 20%
New York	(8.6%, 5.5%)	(10.8%, 6.5%)	New York	(10.9%, 7.2%)	(13.6%, 8.4%)
National	(39.3%, 23.9%)	(49.7%, 27.8%)	National	(51.0%, 32.2%)	(63.9%, 36.9%)
Las Vegas	(83.6%, 49.0%)	(105.6%, 56.4%)	Las Vegas	(115.4%, 69.5%)	(145.5%, 79.1%)

Note: This table shows comparative statics with respect to three important variables: σ , the Poisson probability of a renter becoming a buyer, the REO discount, and the loan balance distribution. The table shows how these variables affect the degree of amplification of the initial shock to prices created by adding foreclosures. The first entry in each pair is the percentage increase in the price index decline generated by defaults over and above that created by the initial shock. The second entry in each pair is the percentage increase in the total sales decline. For instance, when $\sigma = 1/52$ and the REO discount is 20%, the table has an entry of (60.2%, 28.7%) for the 2006 Las Vegas loan balance distribution. Given a price index decline of 4.0% and volume decline of 4.4% from the initial shock alone, this indicates that the full price index and volume declines are respectively 6.4% and 5.7%.

strengthens both the compositional effect and the choosey buyer effects and thus feeds back into further declines in both the overall and non-distressed price indexes.

2.7 Cross-MSA Quantitative Analysis

In order to assess quantitatively the role of foreclosures in the crisis and to test the model's performance, we calibrate the model to national and cross-MSA data described in section 2.2.

Because the size of the preceding bubble is the single best predictor of the size of the ensuing bust, we use a permanent shock to prices that operates through reducing flow utilities a to start the downturn. We assume income shocks last for 5 years, after which they gradually taper off. Recall that there are two parameters that are left to calibrate: σ , which controls how long the average renter takes to return to owner-occupancy, and the size of the permanent shock to prices. To calibrate these parameters, we use the aggregate national data and the cross-MSA data together. We first set a grid of σ s. For each σ , we choose the nationwide permanent shock to prices so that with the nationwide loan balance distribution the model exactly matches the maximum log change in the national house price index. We then simulate the model for each MSA and for each value of σ by assuming that the permanent shock to prices in the MSA is equal to the nationwide permanent shock to prices multiplied by the relative size of the bubble in the MSA as measured by $\frac{\Delta \log(P_{03-06}^{MSA})}{\Delta \log(P_{03-06}^{National})}$. In other words, we assume that the relative amount of housing price appreciation from the bubble that is permanently lost is the same in each MSA.²⁹ We then calculate the

²⁹For instance, the maximum $\Delta \log(P)$ in Las Vegas is was 1.52 times as big as the nation-wide price index. Below we find the permanent price decline for the nation is 21.5%. Thus the permanent decline in Las Vegas is 32.8%.

unweighted sum of squared distances between the maximum change in the log price index in the data and the model for each value of σ and choose the σ with the minimum sum of squared distances.

This methodology yields a sum of squared distances function that is a smooth and convex function of σ . Intuitively too high of a σ will cause the model to over-predict the size of price declines in high-LTV but low-bubble MSAs, while too low of a σ will cause the model to under-predict the size of price declines in high-bubble MSAs. The optimal σ is the one that does best across distribution of bubble size.

The sum of squared distances has a unique minimum, which corresponds to $\frac{1}{\sigma} = 1.05$ years out of the market for the average renter and $\Delta_{\max} P^{National} = 21.5\%$ (a log point decrease of .242).³⁰ A 21.5% price drop from the peak implies that about 2/3 of the price gains between 2003 and 2006 were permanently lost when the bubble burst.

Figure 2.7.1 shows the time series price, sales, the characteristics of distressed sales, probability of sale, foreclosures, and the mass of each type in the market for the resulting national simulation. The qualitative patterns closely match those described in the simpler model in Section 2.4. Recall that there is a one-to-one mapping between the unconditional probability of sale and time to sale, which move in opposite directions.

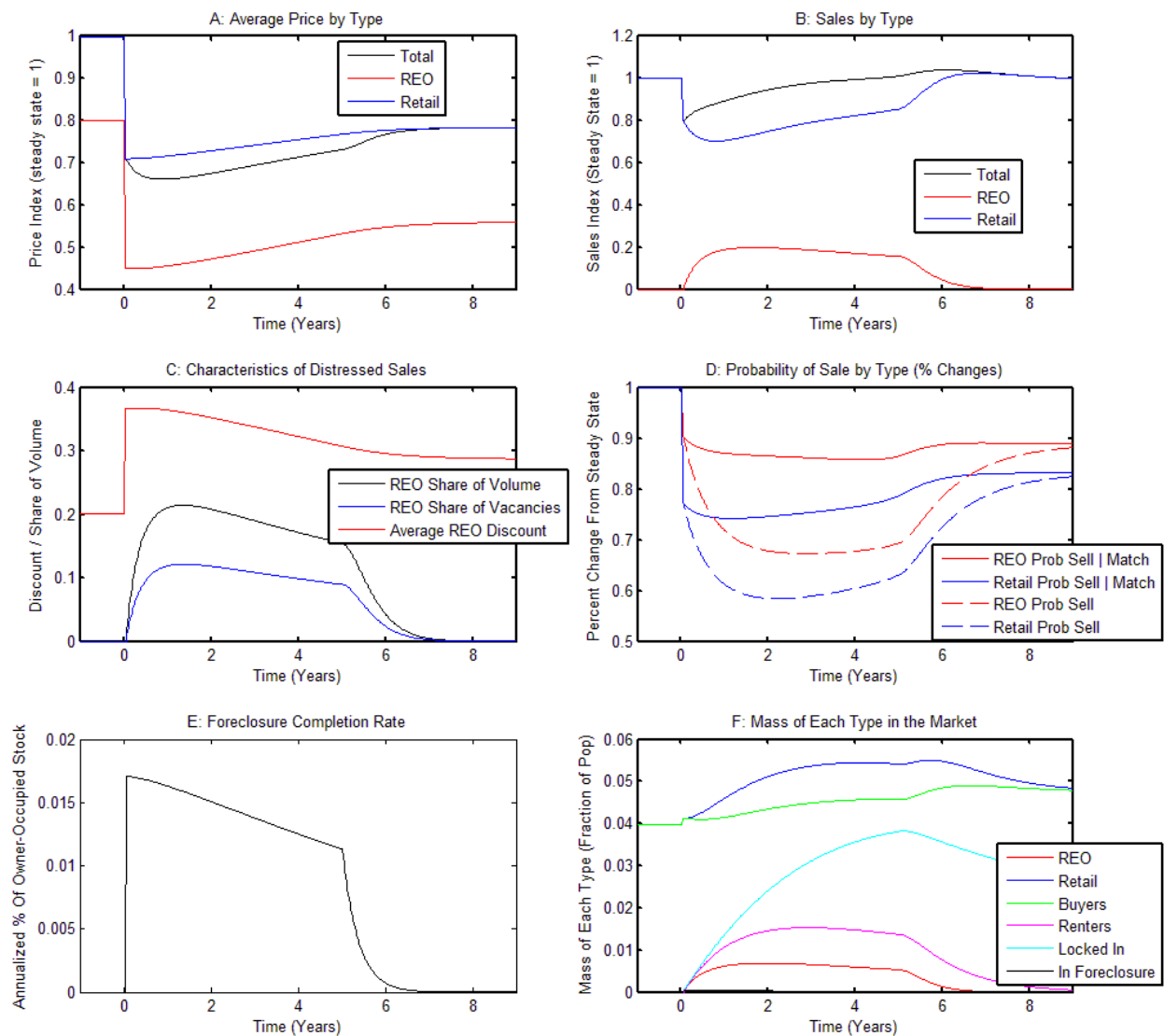
Figure 2.7.2 shows the results of the cross-MSA simulations by plotting the simulated results against the data in six panels that show the maximum change in log price, log retail price, log sales, log retail sales, REO share, and fraction foreclosed. In each figure, the 45-degree line is drawn in to represent a perfect match between the model and the data. The small dots represent MSAs while the large X represents the national calibration.

The calibration procedure does well in matching declines in the aggregate price index across the bubble-size spectrum, as indicated by the fact that the data points are clustered around the 45-degree line. As with the regressions, the extreme outliers are in greater Detroit (the two points in the lower right), which has had a large bust without a preceding boom, and Stockton (the point in the lower middle), which had a much larger bust than boom. Despite these outliers, we cannot reject a coefficient of one when regressing the simulated results on the data, and relative to a case with no default where the entire national price decline is permanent, adding default to the model increases the r-squared of the simulation by 20 percent.³¹ These results suggest that default can explain some of the nonlinearity in Figure 2.2.2.

More importantly, the model cannot be rejected for additional outcomes beyond the aggregate price index that was used for calibration. For the change in log retail price, the national data closely matches the model:

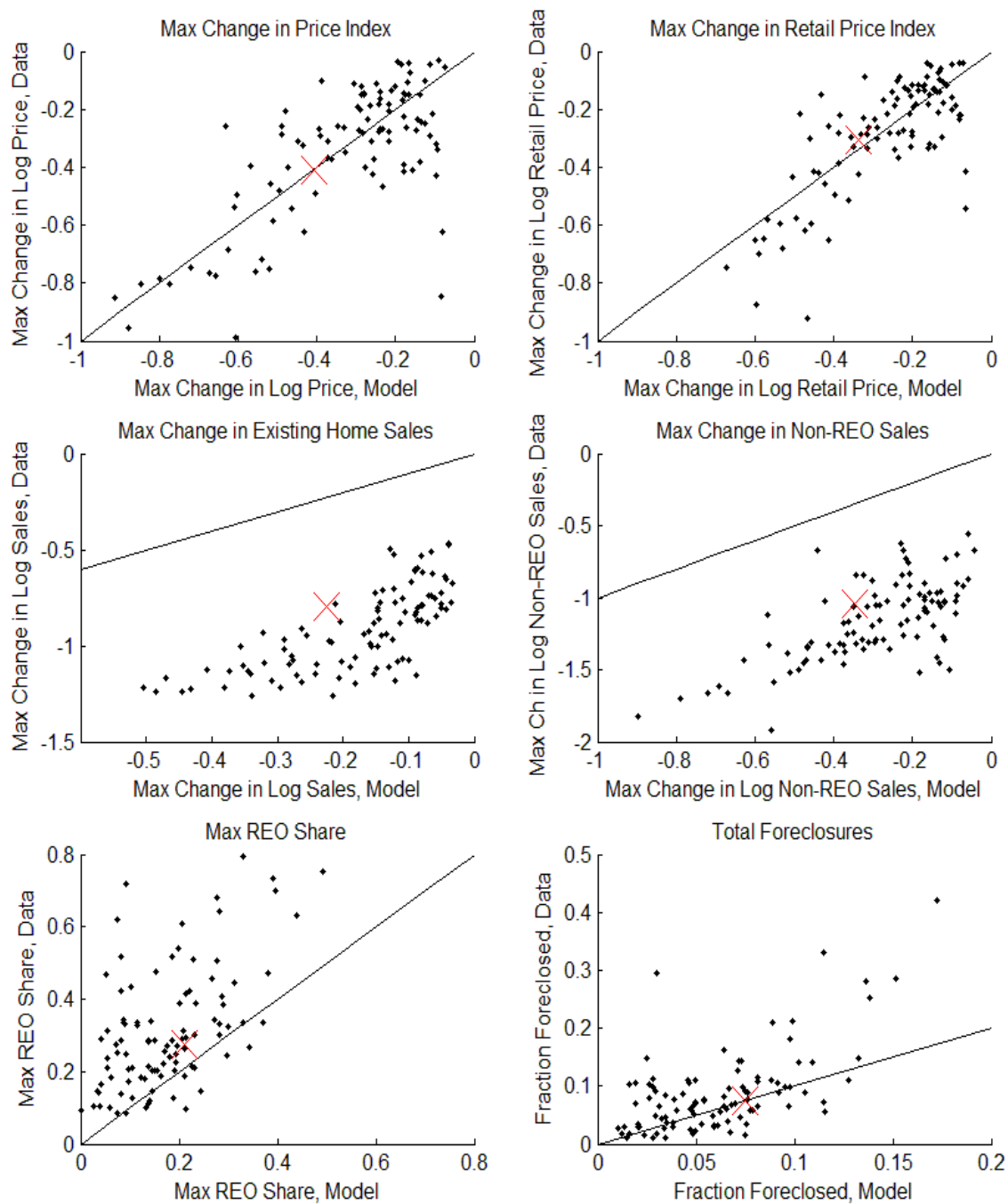
³⁰ 1.05 years out of the market for the average renter may be a bit on the short side, but because we have a constant Poisson probability of leaving renting, the distribution of times out of the market has a very thick tail.

³¹ This r-squared is calculated as 1 minus the the squared distance between the simulated maximum log price decline and its counterpart in the data divided by the squared distance between the data and its mean. This is the r-squared of a regression of the model on the data without a constant.



Note: This figure shows the results of the extended model calibrated to match the national and MSA data as in section 2.7. It uses a permanent shock to housing values and an LTV distribution corresponding to the national housing market. Panels A and B show the average price and sales by type, with pre-downturn price and volume normalized to 1. Panel C shows the REO discount, share of vacancies, and share of volume. Panel D shows the probability of sale conditional on a match and the unconditional probability of sale for each type with the pre-downturn probability normalized to 1. Panel E shows the annualized fraction of the owner occupied housing stock that is foreclosed upon at each point in time. Panel F shows the mass of each type of agent in the market. The overall results are similar to the qualitative results outlined previously.

Figure 2.7.1: National Calibration With Permanent Price Drop



Note: Scatter plots of data vs. simulation results for 97 MSAs in regression analysis. The red X represents the national simulation and each black dot is an MSA. The 45-degree line illustrates a perfect match between the model and the data. The variable being plotted shown in each plot's title. Data is fully described in appendix B.4. The calibration methodology described in text and appendix B.4. The figure shows the model performs well for prices and the number of foreclosures but is off by a constant for volume.

Figure 2.7.2: Cross-MSA Simulations vs. Data

for the most part, the points lie around the 45-degree line, with a few more exceptions where the model under-estimates the change in log retail price such as Stockton and Las Vegas. Nonetheless, a statistical test confirms that we cannot reject that the model matches the data. The model also comes close to matching the data for total foreclosures over 5 years and the model cannot be statistically rejected, although there are some extreme outliers where there were many more foreclosures than the model predicted. Again, some of these are in Greater Detroit and Stockton, but there are a few other hard-hit markets like Las Vegas and the Central Valley in California where the model under-predicts the number of foreclosures. Finally, although it is not in Figure 2.7.2, the national calibration predicts a maximum REO discount of 36.7%. This is slightly above the maximum foreclosure discount for Boston of 35.4% reported by Campbell et al. (2011), so the model can explain time variation in REO discounts.

However, as foreshadowed by the regressions, the model consistently under-predicts the decline in sales. In Figure 2.7.2, the data cluster roughly parallel to the 45-degree line for both retail and total sales, although this is only statistically significant for retail sales. This means that the model does a good job of capturing differences in the size of the maximum sales decline across locations but that volume has fallen nationwide for reasons beyond the model. Potential unmodeled forces reducing volume include the tightening of credit markets, credit constraints and losses on levered properties reducing the purchasing power of buyers (Stein, 1995; Ortalo-Magne and Rady, 2006), a decline in household formation and immigration, a reluctance on the part of retirees to sell their house in a down market, nominal loss aversion (Genesove and Mayer, 2001), increasing returns to scale in matching (Ngai and Tenreyro, 2010), and a reduction in the number of transactions by speculators who flip houses quickly. The cause of the massive decline in volume in the housing downturn is an important subject for future research.

Because it under-estimates the sales decline, the model also under-estimates the REO share in locations that had extremely high amounts of foreclosures, although when we include a constant in a regression of the simulated results on the data we cannot reject a coefficient of one. Because the vast majority of sellers also become buyers, a decline in sales would strengthen the choosey buyer effect, as REO sellers would take up a greater fraction of the market. It would, however, have a much smaller impact on market tightness, because a reduction in the number of buyers and sellers would reduce both the numerator and the denominator. We expect the overall magnitude of the combined general equilibrium effect of foreclosures to be similar.

What do these figures imply about the quantitative extent to which foreclosures exacerbate housing downturns? In the national data, the permanent price decrease that would occur without default is 21.5% (.24 log points) and with default is 33.5% (.41 log points). This implies that the general equilibrium effects of foreclosures together with the compositional effects on the price index induced by a high REO share made the downturn 56% worse than it would have been in the absence of foreclosure. Equivalently, foreclosures

account for 36% of the price decline. This figure is larger in MSAs with larger busts, more default, and a bigger price-default spiral.

The 56% figure, however, includes compositional effects and is thus not the best measure of how much the general equilibrium effects of foreclosure reduce the price a *retail* seller would get if they wanted to sell at the bottom of the market. This is the relevant price for determining negative equity and thus defaults. An alternate metric of the extent to which foreclosures exacerbate downturns, then, is the decline in the retail-only price index, which is 28.7% (.34 in log points) with default and 21.5% without default. The price decline in the retail market is thus 34% worse than it would have been in the absence of foreclosures.

Perhaps surprisingly, these quantitative results are not dramatically changed with an REO discount of 10% in steady state as suggested by Clauretie and Daneshvary (2009) and Zillow (2010). In this case, the same calibration procedure implies a permanent price decline of 22.4% and an average time out of the owner-occupied market for foreclosures of 1.3 years. Intuitively, with the compositional effect weakened by a smaller foreclosure discount, the calibration implies a slightly larger permanent price decline and a stronger market tightness effect. With a 10% steady state discount, the model implies that foreclosures exacerbate the aggregate price decline by 50%. See Appendix B.2.2 for details.

These magnitudes are larger than those implied by other papers. Mian et al.'s (2012) empirical study comes closest to our results. By comparing neighborhoods in states that require judicial approval of foreclosure with neighborhoods just over a border in states that do not, they find that foreclosures were responsible for 20 to 30% of the decline in prices. Our analogous figure of 36% is only slightly higher, likely because we consider market-wide effects that comparing neighborhoods only partially picks up. Calomiris et al. (2008) use a panel VAR to analyze the effect of foreclosures on housing market equilibrium and find that foreclosures would reduce prices by 5.5 percentage points in a foreclosure wave, about half what we find. However, they simulate the impulse response to a wave of foreclosures without a bursting bubble that puts a substantial fraction of homeowners under water, which dramatically increases the size and length of the foreclosure wave. Using a calibrated macro model that focuses on how foreclosures can constrict credit supply, Chatterjee and Eyigungor (2011) find that foreclosures account for 16% of the overall price decline.³²

³²Our results also relate to an empirical literature that examines the effects of REO sales on the sale prices of extremely nearby houses. These papers typically find that a single REO listing reduce the prices of neighboring properties by 1%, with a nonlinear effect (Campbell et al, 2011). The literature is divided as to the mechanism. Anenberg and Kung (2012) arguing that REOs increase supply at an extremely-local level consistent with our market tightness effect, although at a very local level. By contrast, Gerardi et al. (2012) argue that the owners of distressed property reduce investment in their home, effects for which we attempt to control. These papers typically include fine geographic fixed effects and consequently do not pick up the search-market-level effects that are substantial in our model.

2.8 Welfare and Policy Implications

2.8.1 Welfare

To evaluate welfare we adopt a utilitarian social welfare function that equally weights all agents. We can construct social welfare as the discounted sum of individual flow utilities:

$$W = \sum_{t=0}^{\infty} \beta^t \left\{ v_n(t) m_n + v_d(t) m_d + (l_1(t) + f(t)) (h_n^* + 1/\lambda) + \frac{f(t)}{\phi f(t) + 1} c + q(t) \right\}$$

where $q(t)$ follows the law of motion:

$$q(t+1) = (1 - \gamma) q(t) + v_b(t) q_b(\mu(t)) \sum_m r_m(t) (1 - F(h_m(t))) (h_m(t) + 1/\lambda), \quad q(0) = 0.$$

where $q(t)$ denotes the expected flow housing services generated to homeowners in l_0 at time t . We also assume that a foreclosure completion entails certain costs to the bank, such as legal fees and lost revenue from interest payments. A 2008 report by Standard & Poors estimates these costs of foreclosure in excess of the loss on the sale to be approximately \$10,000, so that we set $c = -10$. Finally, we suppose that individuals who receive a taste shock but are unable to move due to negative equity receive no flow utility from their mismatched house.

There are competing effects of foreclosures on social welfare. First, total welfare is decreased relative to the steady state since securing a foreclosure completion is costly, foreclosed homes are sold by REO sellers with higher holding costs, and all homes take longer to sell. More significantly, welfare is decreased by the fact that a number of homeowners receive no flow utility from housing because they are excluded from the housing market for a period of time due to default or locked into a house that does not suit their needs. Second, buyers who do participate in the market are on average purchasing homes which they value more than homes purchased in steady state. Because these buyers stay in these houses for a median of 9 years, this generates a substantial positive effect on welfare that is consistent with anecdotal evidence of the downturn being a “buyer’s market.” Note that the decline in prices which accompanies the downturn has no direct impact on welfare since it operates simply as a transfer from sellers to buyers that has no effect on social welfare. Ultimately, when all of the various effects net out, social welfare falls, but the decline is modest.

This likely understates policy makers’ perception of the social impact of foreclosures. The welfare calculation uses a utilitarian framework and a high discount rate. Given that the downturn is temporary and a number of individuals actually benefit from the housing downturn in the form of increased housing services relative to steady state, it is not surprising we find a modest decline in welfare. However, it is still

the case that a substantial mass of individuals are substantially worse off for several years. To the extent that policy makers adopt a Rawlsian short-term perspective, the social impact of foreclosures could be large.

Most importantly, by focusing only on the housing market, the model misses a number of other potential normative implications of foreclosures. As discussed by Iacoviello (2005), house price declines can have pecuniary externalities because a collapse in home prices destroys wealth in the form of home equity and can impede borrowing by households and firms, creating a financial accelerator effect similar to Kiyotaki and Moore (1997). Moreover, lock-in due to negative equity can impede labor mobility (Ferreira, Gyourko and Tracy, 2010) and exacerbate structural unemployment and can increase the effective risk faced by households since housing consumption is not adjustable (Chetty and Szeidl, 2007). Additionally, banks may be forced to realize substantial losses on foreclosed properties, which impacts their balance sheets through the well-documented net-worth channel in financial intermediation, potentially leading to problems in the interbank repo market, cash hoarding by banks, and a freezing up of credit. Finally, the presence of substantial numbers of foreclosed homes can have negative externalities on communities (Campbell et al., 2011) and can reduce residential investment, construction employment, and consumption (Mian et al, 2011).

While we leave detailed analyses of these important issues to future research, we believe the discussion in this section illustrates there is value in understanding the effectiveness of various policies in ameliorating the foreclosure crisis. We thus conduct a basic positive analysis of three policies that have been proposed that fit into our model: delaying foreclosure, refinancing mortgages at lower interest rates, and reducing principal. To assess the maximum potential impact of each policy, we introduce the policy at time 0. The results of our policy simulations, discussed in the following subsections, are shown in Table 2.8.1. We use the national loan balance distribution and compare the housing market under each policy to a baseline of no policy that is shown in column 1.

2.8.2 Delaying Foreclosure

A simple and low-cost policy that has been proposed is slowing down the pace of foreclosures. To incorporate sluggish foreclosure into our model, we assume that when a homeowner defaults the bank begins foreclosure proceedings but that only $\frac{1}{\phi}$ foreclosures can be processed by the system each week.³³ While in the foreclosure process, it is possible for prices to rise and the house to no longer be in negative equity. If this happens, the foreclosure “cures” and the homeowner lists their house as a normal seller but subsequently become a renter because of the liquidity shock they experienced. In Table 2.8.1, we compare the baseline

³³Formally, we assume that if $f(t)$ homes are in the foreclosure process pending approval only $\frac{f(t)}{\phi f(t)+1}$ can be processed in a given period. We choose this function as a smooth approximation to $\min\left\{f, \frac{1}{\phi}\right\}$, which processes up to $\frac{1}{\phi}$ foreclosures each period. Such an approximation is necessary for the numerical implementation.

Table 2.8.1: Effects of Foreclosure Policies

Policy	Baseline	$\phi = 4000$	$\phi = 6000$	7% \rightarrow 4%	$L \downarrow \$2K$	$L \downarrow \$5K$
$\max \Delta \log (P)$	-.414	-.400	-.387	-.382	-.407	-.396
$\max \Delta \log (P_{\text{Retail}})$	-.343	-.344	-.347	-.327	-.339	-.334
$\max \Delta \log (\text{Sales}_{\text{Existing}})$	-.226	-.205	-.189	-.210	-.223	-.217
$\max \Delta \log (\text{Sales}_{\text{Retail}})$	-.352	-.322	-.295	-.314	-.339	-.321
$\max \frac{\text{Sales}_{\text{REO}}}{\text{Sales}_{\text{Existing}}}$	21.44	17.49	12.61	17.47	20.63	19.41
% Ever Foreclosed	7.55	7.54	7.48	6.10	7.27	6.86

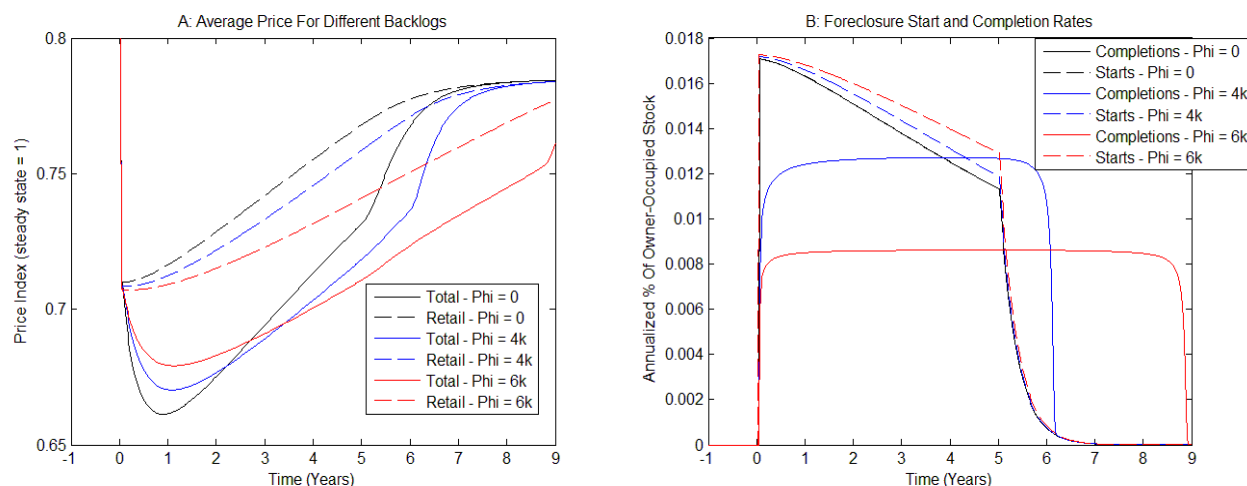
Note: This table shows the impact of various policies on market equilibrium. The first and second rows show the maximum change in log aggregate and retail price, the third and fourth rows show the maximum change in log existing sales and retail sales, the fourth row shows the maximum REO share, and the sixth row shows the fraction of homes foreclosed upon. The first column shows the baseline, which is as in Section 2.7. The second and third columns show the effects of slowing down foreclosures, as $\frac{1}{\phi}$ foreclosures can be processed each week. $\phi = 4000$ and $\phi = 6000$ correspond to a maximum of 1.3 and .9 percent of the housing stock being foreclosed upon per year, respectively. The fourth column shows the effect of reducing interest rates for all homeowners from 7 to 4 percent, a generous estimate of the potential effects of refinancing mortgages. The sixth and seventh columns show the effects of reducing principal by \$2,000 and \$5,000 for all homeowners, corresponding to a \$100 billion principal reduction that is either untargeted or only targeted at under water homeowners.

of $\phi = 0$ to cases when $\phi = 4,000$ and $\phi = 6,000$ so that the maximum annual pace of foreclosure is given by 1.3 percent per year and .9 percent per year, respectively. To further elucidate the effects of foreclosure backlogs, Figure 2.8.1 shows the aggregate and retail price indices and foreclosure starts and completions for the three values of ϕ .

A tighter foreclosure pipeline has different implications for the prices of retail homes versus the overall price index. In particular, the maximum decline in the overall price index falls because the compositional effects of foreclosure are weakened. However, the maximum retail price decline is greater and the price declines last for longer because the foreclosure crisis is extended: as panel B shows, even though foreclosure starts fall off after 5 years, with $\phi = 4,000$ the wave of foreclosures lasts over 6 years and with $\phi = 6,000$ it lasts nearly 9 years. This increases the duration of both the market tightness and choosey buyer effects which gets capitalized into lower retail prices.

The effect of foreclosure backlogs in our model is consistent with the argument that delaying foreclosures does not substantially prevent foreclosures in the long run and only draws out the pain. However, there may be benefits to delaying foreclosure that are not captured by the pure backlog story. For instance, if one expects household formation to pick up and boost demand in the near future, delaying foreclosures from a period of low demand to a period of higher demand could limit price declines. Similarly, slowing down foreclosures could cause banks to offer more mortgage modifications or short sales, reducing the number of delinquencies that result in a foreclosure.

In fact, the empirical evidence on states with judicial approval of foreclosure – in which backlogs are much larger (Mian et al., 2012) – suggests that slowing down foreclosures might reduce the incidence of



Note: The figure shows the effect of prices and foreclosure start and completion rates for three different backlogs. The model is calibrated to the national calibration developed in section 2.7. The pre-downturn price level is normalized to one. One in ϕ houses can be foreclosed upon each week, so $\phi = 4,000$ corresponds to 1.3 percent of the housing stock being foreclosed upon per year and $\phi = 6,000$ corresponds to .9 percent. The figure shows that slowing down foreclosures extends the downturn and makes prices remain low for longer. Although the overall price index rises, this is because of a compositional effect and the retail price index falls.

Figure 2.8.1: Policy: Various Sized Backlogs

foreclosure. Adding a judicial state dummy to regression (2.2.1) leads to a judicial dummy coefficient $+.08$ log points for the aggregate price index and $+.05$ for the non-distressed index even with a full set of controls, as shown in appendix B.3. Our model cannot generate such a dramatic price increase by adding a narrow foreclosure pipeline – the only way to get an effect of this order of magnitude is to reduce the incidence of foreclosures. The welfare effects of policies that limit the ability of lenders to foreclose by slowing down foreclosures are, however, unclear, as lenders may respond to a diminished ability to foreclose by increasing interest rates on mortgages or denying mortgages to credit-worthy borrowers.

2.8.3 Interest Rate or Payment Reductions

Another much-discussed policy that is being implemented with the Home Affordable Refinance Program is to refinance the mortgages of underwater borrowers, many of whom are stuck at extremely high interest rates due to an inability to refinance, at today's low interest rates. This could reduce defaults because some individuals who are currently unable to meet their monthly payment may be able to pay a reduced monthly payment.

To simulate this intervention, we reduce γ_I , the hazard of default for individuals who are underwater, from 8.6 percent to 7.1 percent. Appendix B.4 uses an estimate of the effect of reducing monthly payments

on defaults from Bajari et al. (2010) to show that this reduction in γ_I is equivalent to reducing interest rates from 7 to 4 percent – a generous estimate of what is possible purely through refinancing. The results are shown in column 4 of Table 2.8.1.

Although foreclosures still play an important role in exacerbating the downturn, refinancing has a substantial effect both because mechanically fewer foreclosures occur at a given level of negative equity and because the amplification mechanism is weaker. The size of these effects, however, depends critically on the effect of reducing interest rates on default. While we calibrate to the existing evidence from Bajari et al. (2010), their estimates are not causal. Understanding the impact of interest rate reductions on default is an important subject for future research.

2.8.4 Principal Reduction

The final policy we simulate is a \$100 billion principal reduction.³⁴ The results are shown in Columns 5 and 6 of Table 2.8.1. Column 5 assumes that the government cannot target underwater homeowners gives every mortgage holder a \$2,000 principal reduction. Column 6 assumes that the government can target the approximately 20 million individuals who have had negative equity during the crisis so that principal is reduced by \$5,000 for each underwater homeowner.

Principal reduction provides a direct way to reduce negative equity and the price-default amplification. The targeted principal reduction has a significant ameliorating effect on the crisis, although it is not as effective as the interest rate reduction. The smaller principal reduction, however, has an effect that is much smaller. The government’s ability to target homeowners in need is thus crucial to the effectiveness of principal reduction.

Beyond the government’s ability to target under water homeowners, costs that we do not model may limit the effectiveness of principal reduction. Chief among these is moral hazard: if people expect that underwater mortgages will be bailed out with principal reductions, they may be more likely to become delinquent on their mortgage. Similarly, strategic default may be elevated. The empirical relevance and size of such moral hazard effects is an important subject for future research.

2.8.5 Other Policies

Our policy simulations reveal the trade-offs faced by policy makers and the parameters that future research on anti-foreclosure policy should consider. In addition to the policies simulated here, there are a number

³⁴We assume that raising the funds for the intervention does not affect housing demand, housing prices, or the rate at which liquidity shocks occur.

of other policies that require a richer model to be given full justice and that we hope will be analyzed by future research.

First, a policy maker might try to stimulate additional buyer demand. To have a substantial effect on market tightness, one would have to stimulate entry by new homeowners. Such a policy is outside the scope of our model as it would require endogenous household formation or an endogenous buy-rent decision. Nonetheless, our model does suggest that any increase in new home ownership would have to be permanent; an intervention that boosts new home ownership for a few months at the expense of demand in subsequent months would not have a lasting effect. The short-lived effects of the 2009 new homeowner tax credit suggests that it is difficult to generate a long-lasting effect.

Second, our model cannot consider the conversion of owner-occupied housing to rental housing without an endogenous rent-buy margin. In particular, the conversion of REOs to rental properties has been discussed. While such a policy would reduce rental prices and REO inventories, with endogenous tenure choice it is possible that renting becomes much more appealing, drawing away buyers and further freezing up the owner-occupied market. There is also the potential for rent-seeking behavior by investors who seek to buy REO properties in bulk and convert them to rental homes.

2.9 Conclusion

This paper argues that foreclosures play an important role in exacerbating housing downturns due to their general equilibrium effects. We add foreclosure to a simple search model of the housing market with two types of sellers by making two additional assumptions: banks selling foreclosed homes have higher holding costs than retail sellers and homeowners who are foreclosed upon cannot immediately purchase another home.

With these assumptions, foreclosures alter market behavior by reducing the number of buyers in markets, which makes sellers and particularly REO sellers desperate to sell, and by raising the probability that a buyer meets a REO seller who sells at a discount, which makes buyers more selective. Foreclosures also alter the composition of transactions, making the average sale look more like a foreclosure sale. These effects all create downward pressure on price but have opposing effects on volume as sellers want to sell faster but buyers are more choosy. Sales fall disproportionately in the retail market, helping to explain how foreclosures freeze up the market for non-distressed homes.

We then embed our basic model of the housing market in a richer model which allows for endogenous defaults and homeowner lock-in. We elucidate the potential for spirals in which foreclosures lower prices, putting more homeowners underwater, leading to more defaults and therefore even more foreclosures. A sensitivity analysis demonstrates that such a spiral can operate as a powerful amplification channel of shocks,

especially when the proportion of homeowners in the market with high LTV ratios is high. A calibration of the full model to cross-market data is successful in matching both the average level of the price decline of the housing bust and a significant proportion of the cross-sectional variation in prices. The model matches the cross-sectional pattern of volume declines but is unable to fully account for the level. A quantitative exercise shows that foreclosures exacerbate the price declines in downturns on the order of 50 percent overall and 33 percent in the retail market.

An alternative explanation for the freezing up of the retail market during the housing bust is nominal loss aversion as documented by Genesove and Mayer (2001). The housing bubble may have created a reference point for homeowners such that when the bubble burst, they were not willing to sell for less than what they *perceived* the true value of their homes to be. If this were not the case for banks, the retail market would disproportionately freeze up. However, loss aversion would have to be extreme to explain a freezing up of the retail market for several years. Consequently, while it may not be able to fully account for the freezing up of the retail market, loss aversion may have played a role in the housing downturn and may be able to explain the volume declines that our model cannot capture. Note also that in a model in which nominal loss aversion is an operative channel, foreclosures may actually aid in price discovery.³⁵

Credit constraints and capital losses on levered houses could also explain some of the freezing up of the retail market and the decline in volume that our model cannot explain. Ortalo-Magne and Rady (2006) present a model in which homeowners use equity extracted from their previous house to purchase their next house. With down payment requirements as in Stein (1995), moderate swings in housing prices can generate large swings the purchasing power of potential homeowners. This may cause some homeowners not to move at all, creating effective lock-in of non-under-water borrowers and helping to freeze up the retail market. The substantial decline in household formation is another factor that could explain the decline in volume that our model cannot explain.

Our analysis suggests several directions for future research. First, it would be interesting to endogenize the decision to enter the housing market. Chetty and Szeidl present an (S,s) model of consumption commitments based on the decision to move homes which could be embedded into a general equilibrium model of the housing market. This would allow an analysis of how market forces affect the decision to move and elucidate why sales remain depressed and more people are not taking advantage of the buyer's market by trading up. Second, an endogenous rent-buy margin would allow the rental market to be less segmented and would allow for analyses of several additional policies, as we describe in Section 2.8. Finally, the addition of supply considerations would allow an analysis of how the dynamics of new construction and conversion of

³⁵Thanks to Ed Glaeser for this insight.

owner-occupied housing to renter-occupied are affected by foreclosures.

Chapter 3

Markets with Untraceable Goods of Unknown Quality: Beyond the Small-Country Case¹

3.1 Introduction

In many markets, consumers cannot evaluate the quality of a good before buying it. How can the pleasure of consuming a French brie or a California navel orange be judged without consuming them? How can the feel of a Bic ball-point pen be determined without writing with it? Consumers assess the quality of such goods by purchasing them repeatedly and eventually learn to anticipate their quality.

When the quality of a good cannot be discerned prior to purchase, we call it an “experience good,” a term coined by Nelson (1970). But there are really *two* distinct classes of experience goods. In the first class, the consumer knows the identity of the producer. In the second one, it is either impossible or too costly for the consumer to identify the producer. A Bic ballpoint pen belongs to the first class of experience goods. Behavior in markets where such goods are traded is well understood because of the contributions of Klein and Leffler (1981) and Shapiro (1983) in the IO literature and Falvey (1989) in the trade literature.

This paper focuses on the second class of experience goods. A French brie or California navel belongs to the second class of experience good, where the farm which produced the product is not apparent. Akerlof (1970) initiated investigation of this second class of experience goods but assumed firms could not adjust the

¹Coauthored with Stephen W. Salant and Jason Winfree.

quality of their merchandise. Subsequently, Chiang and Masson (1988) and Donnenfeld and Mayer (1987) in the international literature and Winfree and McCluskey (2005) in the IO literature allowed each firm to adjust the quality of its produce to maximize its profit. As each of these papers clarifies, producers of this second class of experience goods are in a difficult situation. They know the quality of the goods they produce. But they realize that there is no way to distinguish the quality of *their* product from the quality of the other products lumped together in the consumer's mind. They share a "collective reputation." Not surprisingly, a producer does not have as much incentive to make a product of high quality as he would if consumers distinguished his products from those of his competitors.

Each paper in this literature has suggested remedies for this informational externality. Based on their model, Chiang and Masson (1988) propose consolidation of firms in the export sector since firms have stronger incentive to raise quality when the benefit is shared with fewer competitors. Based on their model, Donnenfeld and Mayer (1987) conclude that "a socially optimal policy requires control of both number of export firms and production scale of each firm. One way of affecting the firms' production scale is to set well-defined export quotas for each firm."

While these and other remedies make sense when exporters cannot affect the market price of their products, their value when producers of such experience goods have potential market power remains to be established. In this paper, we depart from Winfree and McCluskey (2005) in allowing firms to choose quantity and from Chiang and Masson (1988) and Donnenfeld and Mayer (1987) in allowing their quantity choice to affect the market price. This enables us to discuss the consequences of alternative remedies for the collective reputation problem when producers have the potential to wield market power.

This case seems relevant to international trade. Over the last decade, China has exported toothpaste, cold medicines, toys, pet food ingredients, dry wall, infant formula and other products laced with lead, antifreeze, and other poisons. Undoubtedly Chinese firms would produce higher quality goods if there were a consolidation of firms as Chiang and Masson recommend and if each remaining firm was subjected to "well-defined export quotas for each firm" as Donnenfeld and Mayer advocate. But their policy recommendations were based on their small country assumption and were never intended for a country like China. In 2007, immediately prior to the Chinese pet food debacle, the Chinese had a substantial 10% share of the total U.S. gluten market. Then it was discovered that Chinese wheat gluten was tainted with melamine. In reaction, U.S. consumers switched to other suppliers, driving up prices on the U.S. gluten market by 30%. In another industry, China managed to raise the real price of cerium (a rare earth) from \$6 a pound in 2008 to \$170 a pound in 2011 by exactly the trade policy (export quotas) which Donnenfeld and Mayer recommend. "The Chinese," as Robert Samuelson noted in a recent column in *The Washington Post*, "do not believe in free trade or fair trade. They practice fixed trade—fixed to benefit them at others' expense." Moreover, he points

out that “by 2030, ...[China’s share of global trade] will reach 15%, twice the American share.”

One potential remedy we consider is a multilateral labeling program which allows consumers to more finely partition all firms operating in the world market for the good. We find that such a program would constitute a Pareto improvement, benefiting all producers and consumers of the good. However, such an effort would likely require considerable international coordination. More realistic policy measures are those undertaken unilaterally by a single country. We find that if a single country undertakes a labeling program for its own firms, such as France has done for its wine producers, the international competitive position of these firms is improved at the expense of firms in other countries. Consumers of the experience good benefit. Similar effects occur if a country unilaterally imposes a minimum quality standard on the firms operating within its borders.

We proceed as follows. Section 3.2 introduces the model and discusses existence issues. In Section 3.3, we investigate remedies to the collective reputation problem. Section 3.4 concludes the paper.

3.2 Model of International Trade

Suppose there are N countries. In country j , n_j firms produce the experience good. The country in which a firm is located is known to consumers. Consumers form a view about the quality of the goods produced in a given country but cannot trace a good to a particular firm within that country.

The players in the game are the $\sum_{j=1}^N n_j$ firms. Firm i in country j sets quality (k_{ij}) and quantity (q_{ij}) simultaneously. Since consumers cannot distinguish firms within country j , every firm in a given country sells its experience good at the same price (P^j) and their merchandise has the same reputed quality (R^j). Given the strategy profile, $\{k_{ij}, q_{ij}\}$ for $i = 1, \dots, n_j$ and $j = 1, \dots, N$, firms in country j develop a reputation for quality equal to the quantity-weighted average of their qualities:

$$R^j = \sum_{i=1}^{n_j} \frac{q_{ij}}{Q^j} k_{ij}, \text{ where } Q^j = \sum_{i=1}^{n_j} q_{ij}. \quad (3.2.1)$$

Assume every consumer gets net utility u from purchasing one unit of the experience good of reputed quality R at price p : $u = a + \theta R - p$. Consumers can purchase a substitute which provides a reservation utility, and they buy the experience good if and only if it provides higher net utility than the outside option. Consumers are assumed to have the same parameters $a > 0, \theta \geq 0$ but to have different reservation utilities.

Consumers observe each country’s reputation for quality. The price they pay depends on the worldwide supply of the experience goods. Suppose that, given the distribution of reservation utilities, a utility of $U(Q)$

must be offered to attract Q customers to the experience good. We assume that $U(Q)$ is strictly increasing, strictly convex, and twice differentiable and that $U(0) = 0$. Price adjusts in each country so consumers are indifferent about the country from which they import the experience good. Every purchaser receives net utility $U(Q)$. Inframarginal buyers strictly prefer the experience good to their outside option while the marginal buyer is indifferent between the experience good and the outside option since both yield net utility $U(Q)$.

As a result, the inverse demand curve is additively separable, differentiable, strictly decreasing and strictly concave: $P^j(Q, R^j) = a + \theta R^j - U(Q)$, for $j = 1, \dots, N$.

We assume that all the output of a given firm has the same quality (k_{ij}) and that each firm has a constant per-unit (and marginal) cost.² The per-unit cost of producing goods of the lowest quality is assumed zero.³ The per-unit cost increases at an increasing rate if the firm chooses to produce higher quality products. In particular, $c(0) = 0, c'(0) = 0$ but $c'(k_{ij}) > 0, c''(k_{ij}) > 0$ for $k_{ij} > 0$.

Consider the game where each firm i ($i = 1, \dots, n_j$) in country j ($j = 1, \dots, N$) simultaneously chooses its output and quality to maximize the following payoff function:

$$q_{ij}[a + \theta R^j - U(q_{ij} + Q_{-ij}) - c(k_{ij})]. \quad (3.2.2)$$

We refer to the second factor in the payoff function as “per unit profit.” Whenever firm i in country j is inactive ($q_{ij} = 0$), its profit is zero.⁴

Since firm i maximizes profits, its pair of decisions must satisfy the following pair of complementary slackness conditions (denoted c.s.):

$$q_{ij} \geq 0, \{a + \theta R^j - U(Q) - c(k_{ij}) - q_{ij}U'(Q)\} + \theta q_{ij} \frac{\partial R^j}{\partial q_{ij}} \leq 0, \text{ c.s.} \quad (3.2.3)$$

$$k_{ik} \geq 0, q_{ij}[\theta \frac{\partial R^j}{\partial k_{ij}} - c'(k_{ij})] \leq 0, \text{ c.s.} \quad (3.2.4)$$

²Since a firm would never choose to offer a portfolio of qualities given the strict convexity of its per-unit cost function, our assumption is unrestrictive.

³Due to concavity, the inverse demand curve will hit the horizontal axis for sufficiently large Q . Price should remain at zero for larger values of Q . That is, technically $\hat{P}^j = \max(0, P^j(Q, R^j))$. There will, therefore, exist trivial Nash equilibria in which every firm sets its quality to zero and Q is large enough that the price in country j is zero. As long as output is distributed among the n_j (for $j = 1, \dots, N$) firms so that no firm could raise price in any country j above zero even if it cut its own output to zero, then every firm would earn zero and no firm could unilaterally deviate to strict advantage. Note that this only works under the assumption that producing nothing at the minimum quality is costless and is not robust to small perturbations in the cost function. Trivial Nash equilibria of this kind also exist in the Cournot model with strictly zero costs. We follow the standard convention of ignoring them.

⁴Since R^j is undefined when $q_{ij} = 0$ and $Q_{-ij} = 0$, the firm's payoff function is undefined at that one point. By defining the firm's profit at that point to be zero, we restore continuity since for $q_{ij} > 0$, profit is zero for any $Q_{-ij} > 0$.

Using (3.2.1), we know $\frac{\partial R^j}{\partial k_{ij}} = q_{ij}/Q^j$ and $\frac{\partial R^j}{\partial q_{ij}} = (k_{ij} - R^j)/Q^j$.

The terms in braces in condition (3.2.3) are standard. They reflect the marginal gain from selling another unit over the marginal cost of producing it. The term following the braces, $\theta q_{ij} \frac{\partial R^j}{\partial q_{ij}}$, is novel and captures an additional consequence (beneficial or adverse) of expanding output marginally. If firm i 's quality is greater than the collective reputation of goods from country j , then increasing output will raise the collective reputation of the good, which will increase the price. On the other hand, if firm i 's quality is below average, then increasing output will lower the collective reputation of goods from country j , which will decrease the price.

The meaning of equation (3.2.4) is straightforward: if the firm produces no output, any quality is optimal; if the firm is active ($q_{ij} > 0$) and quality is set optimally, then a marginal increase in quality must raise the per-unit revenue from sales of the goods by as much as it raises their per-unit cost of production.

When every firm within country j makes the same pair of output and quality choices ($k_{ij} = k_j$ and $q_{ij} = q_j$ for all i in country j), equation (3.2.1) implies $R^j = k_j$, $\frac{\partial R^j}{\partial q_{ij}} = 0$, and $\frac{\partial R^j}{\partial k_{ij}} = 1/n_j$. Hence, the following conditions must be satisfied by the $2N+1$ variables (Q and $\{q_j, k_j\}_{j=1}^N$) in a symmetric Nash equilibrium:

$$q_j \geq 0, \quad a + \theta k_j - U(Q) - c(k_j) - q_j U'(Q) \leq 0, \quad \text{c.s.} \quad (3.2.5)$$

$$k_j \geq 0, \quad q_j \left[\theta \frac{1}{n_j} - c'(k_j) \right] \leq 0, \quad \text{c.s.} \quad (3.2.6)$$

and

$$Q = \sum_{j=1}^N n_j q_j. \quad (3.2.7)$$

The following theorem proves that a Nash equilibrium must be symmetric and that there exists a single solution to these $2N + 1$ necessary conditions.⁵

Theorem 3.2.1 *If there exists a Nash equilibrium, it is unique and every firm within a given country produces the same quantity at the same quality.*

Proof. In the appendix, we exclude the possibility of any Nash equilibria where firms within some country choose different quantities or qualities. To exclude the possibility of *multiple* Nash equilibrium where firms within each country behave the same, we show there exists a unique symmetric solution to the conditions

⁵In footnote 2, we identified equilibria resulting in a zero price which we exclude as trivial. We also exclude solutions to (5)-(7) that arise because quality, although the same for any strictly positive output of the firm, is indeterminate when production is exactly zero. We do so by requiring that the firm choose the same quality when $q_j = 0$ as it would when producing strictly positive output. In the absence of such a requirement, a firm could produce nothing at a quality so high that the cost of producing anything positive would exceed the revenue which could be gained, thereby satisfying the necessary conditions. We exclude such solutions. But even if we had included them as solutions, they can be ruled out as Nash equilibria.

necessary for a Nash equilibrium. If firms in country j are active ($q_j > 0$) then condition (3.2.6) can be satisfied only by the unique quality choice making the factor in square brackets zero. If firms in country j are *inactive*, we assume that the firm chooses this same quality (see footnote 4). So we substitute $k_j = c'^{-1}(\theta/n_j) > 0$, for $j = 1, \dots, N$ in the N equations of (3.2.5) and seek to solve (3.2.5) and (3.2.6).

Replace Q by X in condition (3.2.5). The unique solution to this condition can be written as the continuous function $q_j(X) \geq 0$. This function has a strictly positive intercept ($q_j(0) > 0$), strictly decreases until it reaches zero and remains zero for larger X .

Consider next $\sum_{j=1,N} n_j q_j(X)$. This function of X must also be continuous with a strictly positive intercept. It must also strictly decline until it reaches zero and then must remain zero for larger X . It follows that $\sum_{j=1,N} n_j q_j(X)$ must cross the 45 degree line exactly once. At the unique crossing, $\sum_{j=1,N} n_j q_j(X) = X > 0$. Denote this unique fixed point $Q > 0$. By construction, given our solution of (3.2.6) there is a unique solution to (3.2.5) and (3.2.7). ■

Therefore, the only candidate for a Nash equilibrium is the unique fixed point in which all firms within a country make the same quality and quantity choices. We show now that:

Theorem 3.2.2 *At any Nash equilibrium, all firms are active.*

Proof. We confine attention here to the unique equilibrium, if it exists, in which firms within each country are symmetric since the appendix establishes that there can be no other Nash equilibria. To show here that all firms must be active in this equilibrium, we establish that whenever firms in some country j are inactive ($q_j = 0$), any of them could unilaterally deviate to strict advantage; hence the unique solution of the first-order conditions does not constitute a Nash equilibrium. Suppose in the unique fixed point that $q_j(Q) = 0$ and $a + \theta k_j - U(Q) - c(k_j) - q_j U'(Q) \leq 0$ in some country j . We can rule out this solution as a symmetric Nash equilibrium by the following argument. If this condition were to hold, then since it is strictly decreasing in q_j , there can be no positive quantity choice that can satisfy the condition with equality for that same $k_j = c'^{-1}(\theta/n_j) > 0$. Hence, a firm's best reply to the other players' strategy profile would be zero production. But a firm in country j can strictly improve on that. Since $Q > 0$ there must be some country h with strictly positive output and hence with $a + \theta k_h - U(Q) - c(k_h) = q_h U'(Q) > 0$. A firm in country j could choose the same quality k_h , marginally increase its output above zero, and make a strict profit. Hence, the premise that there can be a symmetric Nash equilibrium with $q_j(Q) = 0$ must be false. ■

It follows that there can be at most one Nash equilibrium and in it every firm within a country behaves in the same way and is active ($q_{ij} = q_j > 0$ for all i, j). Even if every firm is active in the unique fixed point,

firms in some country may be able to strictly improve their profits with a *nonlocal*, unilateral deviation.⁶ However, a unique Nash equilibrium always exists for sufficiently small θ . This follows since when $\theta = 0$ the model collapses to the Cournot model with strictly concave inverse demand $P^j = a - U(Q)$. Since consumers do not value quality in this extreme case, firms would not incur the expense of providing it and the per unit cost of expanding output would be zero. Such a Cournot model is known to have a unique nontrivial equilibrium. By continuity it can be established that our model will continue to have a unique Nash equilibrium provided $\theta > 0$ is sufficiently small.

In characterizing the nature of this equilibrium, it turns out that the degree of competition *within* a country is of central importance, even though all firms compete on the world market and there is a single inverse demand curve faced by all firms.

Theorem 3.2.3 *Countries with a larger number of firms export lower-quality goods. Every firm within such countries exports less and earns lower profits than in countries with a smaller number of firms.*

Proof. Equation (3.2.6), which holds for firms in each country j , implies that a country with a larger number of firms exporting the experience good (the “larger country”) will export lower-quality goods. Since prices adjust so that consumers are indifferent about the source of their imports, the exports of larger countries must sell for lower prices.

In the equilibrium, any firm offering quality k earns profit per unit of $a + \theta k - U(Q) - c(k)$. Since this function is strictly concave in quality and peaks at k^* , the implicit solution to $\theta = c'(k)$, in equilibrium every firm will choose a quality $k_j < k^*$. Therefore, the profit per unit at each firm in a group rises if the common quality of every firm in that group increases. It follows that firms in larger countries will have lower profit per unit. But equation (3.2.5) implies that any firm with a lower profit per unit produces less output and hence earns lower total profit.⁷ ■

This result is intuitive. Since the quality reputation of an individual firm within a country is linked to the quality choices of its competitors in that *same* country, there exists a classic free-riding problem in which an individual firm has little incentive to invest in quality provision. A greater number of firms within a country exacerbates the severity of this collective reputation problem because the perceived marginal benefit of quality investment is inversely proportional to the degree of competition within a country. A lower collective

⁶Since the solution satisfies the necessary conditions, no local unilateral deviation will be profitable.

⁷In this model, it is assumed that firms producing in one country cannot disguise their products as originating in another country where firms have a better reputation and earn higher profits. Presumably this implicitly requires that the government identify and prohibit such deceptions since they would be profitable. “Under EU law, for example, use of the word Champagne on wine labels is intended exclusively for wines produced in the Champagne region of France under the strict regulations of the region’s Appellation of Controlled Origin . . . Customs agents and border patrols throughout Europe have seized and destroyed thousands of bottles in the last four years illegally bearing the Champagne name, including product from the United States, Argentina, Russia, Armenia, Brazil and Ethiopia.” Castillo (2008)

reputation makes that country's firms less competitive on the world market and therefore less profitable. In equilibrium, less profitable firms produce less and earn less gross profits. In the special case where every country has the same number of firms, quality, price, output, and profit are the same at every firm regardless of its location.

3.3 Potential Remedies

We now investigate potential remedies to the collective reputation problem. These can be divided into remedies that require international coordination and remedies that a country can pursue unilaterally. Given that consumers observe only the country of origin of the experience good but not the identity of the firm which produced it, we investigate the effects of giving each consumer the information with which to classify more finely the source of the imported good. Consumers may be better able to better classify every firm in the world or merely better able to identify the firms from a single country. We also consider the qualitative impacts of a minimum quality standard on exports imposed by one country on firms within its borders.

3.3.1 Multilateral Remedies

Suppose consumers receive sufficient information to partition more finely the source of the imported good, regardless of its country of origin. Specifically, we have the following result:

Theorem 3.3.1 *Let \mathfrak{N} denote the worldwide partition of firms.⁸ Suppose a labeling program permits consumers to classify firms into a strict refinement $\mathfrak{N}' \subset \mathfrak{N}$ so that every firm is now assigned to a smaller group. Then quality, quantity, and profit will increase at every firm and the utility of every consumer of the experience good will increase as well.*

Proof. If every firm is a member of a smaller group, world production must increase. For, suppose it decreased. Then fewer consumers would buy the good, and it must provide lower utility. But since each firm shares its collective reputation with fewer competitors, every firm will increase its quality, and hence the reputed quality of its group will increase. But since profit per unit is increasing in the common quality of the group, the sum of the first four terms in equation (3.2.5) must increase. Since the second factor of the last term decreases, however, that equation would hold only if each firm's output (q_j) increased. But then world output would increase, contradicting our hypothesis.

Consequently, world sales of the experience good must strictly increase and hence so must production at each firm. Since each firm belongs to a smaller group, each firm will increase its quality. As for profitability,

⁸In particular, \mathfrak{N} has N subsets with the j th subset consisting of n_j firms.

since both factors in the last term of equation (3.2.5) increase, net profit per unit (the first four terms of that equation) must increase. Therefore profit at every firm will rise. To absorb the increased production, the utility each consumer receives from the experience good must increase by enough to attract the requisite number of consumers away from their outside options. ■

Profits, and hence social welfare, would be higher if every consumer recognized not only that a product was imported from a particular country but also that it was made in a particular region of that country (or, better yet, by a particular ethnic group within that region). For, the smaller the number of firms in a category recognized by every consumer, the less incentive each firm will have to shirk in the provision of quality.⁹

3.3.2 Unilateral Remedies

Coordination problems and political differences might make a worldwide labeling program difficult to implement. Instead a government might want to classify firms within its own borders unilaterally or alternatively to impose a minimum quality standard on the experience good its firms export. We will assume that such a standard (\bar{k}) is binding but is not set as high as the one a firm would choose if it were the only domestic producer. That is, we assume $\bar{k} < k^*$, where k^* solves $\theta = c'^*$. We establish that:

Theorem 3.3.2 *A finer classification of firms within a single country or the imposition of a minimum quality standard by one country raises the output and profits of its exporters while lowering the output and profits of unregulated exporters elsewhere. Overall, world output expands. Quality rises in the country in which the program originates and remains unchanged elsewhere. Both programs benefit every consumer of the experience good.*

Proof. The imposition of either the labeling program or the minimum quality standard by a single country i must strictly increase world production of the experience good. For, suppose the contrary. Suppose aggregate quantity falls or remains constant. Then the utility which consumers get from the experience good must weakly decrease. In every country $j \neq i$, exporters would maintain quality since equation (3.2.6) still holds. So if their exports provide weakly less net utility, the prices of their exports ($P^j = a + \theta k_j - U(Q)$) must weakly increase. Since the per unit profit ($P^j - c(k_j)$) would then weakly increase, equation (3.2.5) implies that output at each such firm must weakly increase.

As for the firms in country i , their per unit profit must *strictly* increase. If the country is able to more finely classify its firms on the world market, then each firm is a member of a smaller group and therefore

⁹In the limit of perfect firm traceability, each firm produces the socially optimal quality due to additive separability of the inverse demand function. Quantity choices remain suboptimal due to the usual oligopoly considerations.

produces a higher quality by equation (3.2.6). Alternatively, quality is raised at each firm by force of the binding minimum quality standard which, by assumption, was not excessive ($\bar{k} < k^*$) and therefore increases per unit profit. Equation (3.2.5) then implies that output at each firm in country i strictly increases. But then we have a contradiction: aggregate output cannot weakly decrease as we hypothesized since that implies the sum of the individual firm outputs would strictly increase.

So the imposition of either a labeling program or a minimum quality standard by a single country i must cause world output of the experience good to strictly increase and hence must cause the net utility of every consumer of the good to increase. Since the quality of the firms in countries $j \neq i$ does not change, their prices, profit per unit, output and total profits must fall. Since aggregate output expands despite the contraction at every such firm, output at every firm in country i must increase. But, as equation (3.2.5) implies, such firms would expand output only if their profit per unit also increased. Hence, their total profits would also increase. Since profit per unit increases at each regulated firm, its price per unit must increase by more than enough to offset the increased cost per unit of producing the higher quality. ■

Intuitively, both programs create a competitive advantage for the firms within the country implementing it and thus, *ceteris paribus*, disadvantages firms located in other countries. The labeling program alleviates the free-riding problem created by the collective reputation and the minimum quality standard reassures buyers about the quality and safety of products originating from that country. In effect, the minimum quality standard eliminates the strategic difficulties caused by collective reputation by force of regulation.

3.4 Conclusion

In this paper, we considered markets where consumers cannot discern a product's quality prior to purchase and can never identify the firm which produced the good. We began by considering an international model in which the information set of consumers is the country of origin and the average quality of the experience good in that country. In the absence of any regulations, a country with a larger number of producers of the experience good will export shoddier products at lower prices. When buyers who care about the quality and safety of the products they purchase are able to classify producers into even finer partitions, firms are motivated to improve the quality and safety of their merchandise. As a result, both profits and consumer welfare increase. A minimum quality standard can secure further benefits. If one country imposes such regulations, consumers benefit not only from the enhanced quality of that country's exports but from the opportunity to buy other countries' exports which sell for diminished prices despite their unchanged quality. The minimum quality standard imposed by one country raises the profits of the firms compelled to obey them and reduces the profits of competing exporters with the misfortune to be located in countries without

such regulations.

Although we have focused here on the implications of collective reputation for international trade, the issue of collective reputation also has implications for domestic antitrust policy. The Department of Agriculture (USDA) has long advocated minimum quality standards for fruit and vegetables in recognition of the collective reputation problem which farmers face. The Department of Justice, however, regards all such minimum quality standards as mere volume restrictions intended to benefit farmers at the expense of the consumers. For example, DOJ (Bingaman and Litan, 1993) objected to the minimum quality standards that USDA advocated for oranges, grapefruit, tangerines, and tangelos grown in Florida, for tart cherries grown in Michigan and for oranges and grapefruits grown in Texas. In future work, we hope to identify circumstances where such minimum quality standards would be welfare-improving and should be allowed.

Chapter 4

Regulating Experience Goods in Developing Countries¹

4.1 Introduction

In developing countries, informal markets co-exist alongside formal markets. Contrary to the conventional view that the informal sector produces intermediate goods for the formal sector, in many cases, the same kinds of goods and services are provided by both sectors and are in fact manufactured by similar processes. Empirical studies (e.g. Myint (1985) and Livingstone (1991)) have documented a “quality gap” between the outputs of the two sectors. Goods produced in the formal sector are, on average, of higher quality and more expensive than those produced in the informal sector.

Banerji and Jain (2007) have recently provided one possible explanation for this “quality dualism.” In their model, the formal sector is characterized by higher wages and lower rentals and firms decide which sector to be in. Given this exogenous difference in factor prices and the assumption that higher quality goods are more capital-intensive, the formal sector firms can make high quality products more cheaply and the informal sector firms can make low quality products more cheaply. Since consumers differ in how much they are willing to pay for better quality, both sectors have a demand for their products.

However, Banerji and Jain cannot explain much of the quality dualism in developing countries since their explanation requires that consumers ascertain quality prior to purchase. In this paper, we explain how a quality gap can arise when their assumption fails to hold. The quality of some goods can be determined by a simple inspection prior to purchase, but in many other cases, the quality of a good can be determined only

¹Coauthored with Stephen W. Salant and Jason Winfree.

after consuming it. Such “experience goods” can be classified into two categories: in the first, a consumer can costlessly identify the producer of the good; in the second, the consumer cannot at reasonable cost identify its producer.² It is the latter which we consider in this paper.

Even in developed countries, this second class of consumer goods is not uncommon. When we buy a Washington apple, we do not know how it will taste and we cannot ascertain what orchard produced it. Moreover, globalization has raised the frequency with which consumers in developed countries encounter such products. But the inability to trace an experience good to its producer is even more common in developing countries regardless of whether the good is sold in the formal or informal market.

Consider some examples. Fruits, vegetables, meats, fish, grains, coffee beans, videos, and other goods are sold both in formal stores and on the street in the informal market. The quality (both the taste and the health consequences) of consuming these goods is uncertain prior to purchase; often the purchase must be made without knowing who grew or manufactured the product. Cooking oil in India is sold both in groceries and on the street. In either case, a consumer does not know whether it has been adulterated, nor is the source of the oil easy to determine. The same is true of maize in Kenya. A consumer can buy it in a grocery or on the street. But in either case, there is uncertainty about whether it has been stored improperly and become contaminated with aflatoxin, a fungus that cannot be detected by casual visual inspection. Panela, a sugar produced in massive quantities in Colombia and elsewhere in Latin America, is another example of a good sold in both formal and informal markets. In each case, the quality is impossible to ascertain prior to purchase, and any of a large set of households or firms might have produced it.

When the consumer cannot identify the producer of an experience good prior to purchase, he must base his expectation of the product’s quality on something else. We assume here that he *can identify* the sector which produced the good. Knowing the quantity of goods of each quality produced by that sector, he anticipates that the quality of the good he is considering equals the quantity-weighted average quality of goods from that sector.

This informational environment puts producers of such goods in a difficult position. They know the quality of the goods they produce. But they realize that there is no way for a potential buyer to distinguish the quality of *their* product from the quality of the other products lumped together in the consumer’s mind. They recognize, therefore, that they share a “collective reputation.” Not surprisingly, a producer does not have as much incentive to make a product of high quality as he would if consumers distinguished his products from those of his competitors.³

²There are obviously also intermediate cases.

³This is different than the models of Shapiro (1983) and Klein and Leffler (1981) in which consumers cannot ascertain quality prior to purchase, but can at least identify the producer of the experience good by scrutinizing its branding or packaging. They

In this paper, we consider n producers of a single experience good. While in the short run firms are exogenously assigned to the formal or the informal sector, in the long run firms are free to costlessly move between the two. Producers operating in the formal sector always sell the good to consumers through perfectly competitive outlets in the formal market, while producers in the informal sector sell the good through outlets in the informal market. Consumers may patronize either market (or neither market).

One issue which we have heretofore neglected is what we mean by “formal” and “informal.” Since these terms have not always been used consistently in the literature, it is important to be precise. Some analysts distinguish the two sectors by endogenous market variables, as in Banerji and Jain (2007). Just why the sectors display this difference in factor prices is outside their formal model, but they attribute it to government regulation. Other analysts *center* the distinction between the two sectors on regulation.⁴ Mazumdar (1983) defines formal labor markets as markets that have labor laws or unions. Marjit, Ghosh and Biswas (2007) assume that informal markets imply illegal production. In Rauch (1991) minimum wage regulations are imposed only on firms of above a certain size.

In our paper, producers which submit to government regulation are said to be in the formal sector while firms which fly under the regulatory radar are said to be in the informal sector. If the government removed every regulation from formal sector producers, then, in the long run as firms re-allocate themselves, the sectors would be of equal size and every firm would produce goods of the *same* quality regardless of the sector it was in, and these goods would sell for the *same* price. This is consistent with the observation of Castells and Portes (1989, p. 13) that “with no regulation of any kind, the distinction between formal and informal would lose meaning.”⁵

Introducing a regulatory structure creates a meaningful distinction between the formal and informal sector and can endogenously lead to a quality gap in the output of the two sectors. One simple policy instrument which generates quality dualism and which has received considerable attention in the development literature is taxation.⁶ In our model, as taxes are levied on formal sector producers, the long run equilibrium is characterized by an informal sector which is larger than the formal sector and produces lower quality, lower priced goods. This is because the introduction of taxes leads firms to migrate from the formal sector to the informal sector in order to escape the tax burden. This leads to an expansion of the informal sector and

therefore base their expectation of the quality of the experience good on the reputation of the firm which produced it. Given that the consumer can identify the producer of the experience good, one firm’s quality choice *never affects* another firm’s reputation for quality.

⁴For an overview of the differences between formal and informal see Dixit (2004). Some of the early discussions of informal markets come from the International Labour Organisation (1973) and Mazumdar (1983).

⁵Quality provision in each sector would still be suboptimal from a social welfare perspective due to the freeriding problem caused by the collective reputation.

⁶See Fortin et al. (1997), Dessy and Pallage (2003), Ihrig and Moe (2004).

a contraction of the formal sector, thereby exacerbating the collective reputation problem in the informal sector, while ameliorating it in the formal sector. As a result, quality decreases in the informal sector and increases in the formal sector. Since consumers can patronize either market and require lower prices for goods expected to be of lower quality, informal market goods must sell for less than formal market goods.

Note that regulation of the formal sector will affect firm behavior in the informal sector. The economics literature on the formal and informal sectors has long recognized that regulating the formal sector will have such spillover effects. Most recently, for example, Marjit and Kar (forthcoming) consider a Heckscher-Ohlin-Samuelson (HOS) model with each good produced in both the formal and informal sectors. Each good is of known and invariant quality. Marjit and Kar show that a decline in tariff rate should increase the informal wage in a developing economy.

Not all government regulation would lead firms in the formal sector to produce higher quality goods. For example, a subsidy per unit produced in the formal sector would lead to a long-run equilibrium with a larger formal sector and with the informal sector producing the higher quality goods. A binding minimum quality standard imposed exclusively on firms in the formal sector eliminates the collective reputation of firms in that sector and will attract firms from the informal sector. In the long run equilibrium all firms, including those *outside* the ambit of regulation (i.e. firms in the informal sector), produce at the quality level specified by the standard. Hence, the government could use regulation to *eliminate* the quality gap and to achieve socially optimal quality provision in each sector.

In the next section, we introduce our model. In section 4.3 we discuss the effects of regulation on the two sectors in the short and long run. Section 4.4 concludes.

4.2 Short and Long-run Equilibrium in the Absence of Regulation

We consider a country in which a single experience good is produced in both a formal (F) and an informal (I) sector. The formal sector is visible to the government and is subject to regulation. Conversely, the informal sector consists of producers that operate beyond government scrutiny and are therefore not subject to regulation. We suppose that there are n producers of the final good, which operate in at most one sector. For simplicity, we do not require that n be an integer. In the short run, producers are assigned exogenously to sectors; in the long run, the producers migrate to the sector which is more profitable.⁷ Each producer, denoted by i , in sector j for $j \in \{F, I\}$ sets quality $k_{ij} \in [0, \infty)$ and quantity $q_{ij} \in [0, \infty)$ simultaneously to maximize profits given the conjectured behavior of the other firms and the anticipated decision rules of the

⁷We make this assumption only for tractability. Considering measure-theoretic analyses where firms are infinitesimal in size changes none of the qualitative results and serves only to add unimportant technical considerations.

consumers.

A consumer gets net utility u from purchasing one unit of quality k at price p : $u = \theta k - p$. Each consumer has an outside option \bar{u} and buys the experience good if and only if it provides higher net utility than his outside option. We assume that consumers have the same θ but different outside options.⁸

Consumers purchase the experience good from perfectly competitive sellers operating in either the formal market or informal market. We assume that sellers in the formal market obtain the good only from producers in the formal sector, while sellers in the informal market obtain it only from producers in the informal sector. Consumers form independent views about the quality of the goods being sold in the formal and informal markets but cannot trace a good to a particular producer. Instead, consumers only know the distribution of quality among producers in each sector. Since consumers are assumed to be risk neutral, it therefore follows that the expected utility of purchasing the experience good in either the formal or informal market at price p is given by $u = \theta R^j - p$, where:

$$R^j = \sum_{i=1}^{n_j} \frac{q_{ij}}{Q^j} k_{ij} \quad (4.2.1)$$

Here n_j is the number of producers in sector j and Q^j is the total quantity of the good produced by sector j . In other words, every producer in a particular sector shares the same collective reputation for quality equal to the quantity-weighted average of qualities in that sector. Because consumers cannot trace an individual good to its producer, there is a single price P^F in the formal market and a single price P^I in the informal market.

Consumers observe the average qualities and prices of goods sold in the formal and informal markets, respectively. Shoppers compare $\theta R^F - P^F$, $\theta R^I - P^I$, and \bar{u} , and patronize the source providing the highest net expected utility. In the event of ties, consumers may patronize multiple sources. In equilibrium, prices adjust so that consumers are indifferent between purchasing the experience good on either the formal market or informal market but may strictly prefer both to their outside options. Prices depend on the aggregate supply of the experience good produced in the two sectors. Suppose that, for the given distribution of outside options, a utility of G must be offered to attract Q customers to the experience good. We assume that $G(Q)$ is a strictly increasing, strictly convex, and twice differentiable function and that $G(0) = 0$. $G(Q)$ reflects the distribution of the reservation utilities of the heterogeneous consumers.⁹ In equilibrium, everyone purchasing the experience good on either the formal or informal market receives the same net expected utility $G(Q)$. Some patrons (the inframarginal buyers) strictly prefer the experience good to their outside option; other

⁸If the experience good is a “consumable” such as food, the outside option could represent home production. The heterogeneity would then reflect differences in gardening skill and in the time at home which can be devoted to home production.

⁹Formally, let μ be a σ -finite measure on $[0, \infty)$. Define $Q(G) = \int_0^G d\mu$. We assume that $Q(0) = 0$ and that $Q(G)$ is twice differentiable, strictly increasing, and strictly concave. Let $G(Q) \equiv Q^{-1}(G)$.

patrons (the marginal buyers) are indifferent between the experience good and the outside option, both of which yield net utility $G(Q)$.

More formally, let the inverse demand of sector j ($j = F, I$) be given by:

$$P^j(Q, R^j) = \begin{cases} \theta R^j - G(Q); & \text{if } Q < G^{-1}(\theta R^j) \\ 0; & \text{if } Q \geq G^{-1}(\theta R^j) \end{cases} \quad (4.2.2)$$

The inverse demand curves for the goods produced in each sector is strictly increasing in that sector's reputed quality, strictly decreasing and strictly concave in the total output of the two sectors, and additively separable in the two variables over the set $\{(Q, R^j) | Q < G^{-1}(\theta R^j)\}$.

Suppose *ex-ante* there are n_F producers registered in the formal sector and n_I producers operating in the informal sector. Also, for the moment consider the benchmark with no regulations. Consider the game where each producer i ($i = 1, \dots, n_j$) in sector j ($j = F, I$) simultaneously chooses its output and quality to maximize the following payoff function:

$$q_{ij}[P(q_{ij} + Q_{-ij}, R^j) - c(k_{ij})] \quad (4.2.3)$$

where $c(k_{ij})$ is the strictly increasing, convex per unit cost if output is of quality $k_{ij} > 0$; in addition, $c(0) = c'(0) = 0$.

When the other firms in sector j are producing ($Q_{-ij} > 0$), then the reputed quality of sector j is well defined whether or not firm i is producing. Hence, equation (4.2.3) is well defined. When no firm in sector j is producing, however, the reputed quality of that sector is ambiguous, and we *define* the profit of firm i , given in equation (4.2.3), as zero. This assignment never conflicts with (4.2.3), as that equation, evaluated at $q_{ij} = 0$, gives the same result when $Q_{-ij} > 0$ and is undefined when $Q_{-ij} = 0$.

Because firm i maximizes profits, its decisions must satisfy the following pair of complementary slackness (denoted c.s.) conditions for $Q_{-ij} > 0$:

$$q_{ij} \geq 0, \quad \theta R^j - G(Q) - c(k_{ij}) - q_{ij}G'(Q) + q_{ij}\theta \frac{\partial R^j}{\partial q_{ij}} \leq 0, \quad \text{c.s.} \quad (4.2.4)$$

$$k_{ij} \geq 0, \quad q_{ij}[\theta \frac{\partial R^j}{\partial k_{ij}} - c'(k_{ij})] \leq 0, \quad \text{c.s.} \quad (4.2.5)$$

where, from (4.2.1), $\frac{\partial R^j}{\partial k_{ij}} = q_{ij}/Q^j$ and $\frac{\partial R^j}{\partial q_{ij}} = (k_{ij} - R^j)/Q^j$. In the appendix, we show that there exist no Nash equilibria of the game where firms within a sector use differing strategies. We also establish that conditions (4.2.4) and (4.2.5), which must hold at any Nash equilibrium, have only one solution where

firms within a sector follow symmetric strategies and receive a strictly positive price. This unique solution constitutes a Nash equilibrium whenever all firms are active and none can make a nonlocal, profitable deviation. In the Nash equilibrium, every firm in a sector j sells the same unique, strictly positive amount ($q_{ij} = q_j > 0$, for all i) at the same unique, strictly positive quality ($k_{ij} = k_j > 0$, for all i). It follows that $\frac{dR^j}{dk_{ij}} = \frac{1}{n_j}$ and $\frac{dR^j}{dq_{ij}} = 0$.

We can rewrite the first-order conditions (4.2.4) and (4.2.5) as follows:

$$\theta R^j - G(Q) - c(k_j) - q_j G'(Q) = 0 \quad (4.2.6)$$

and

$$\theta \frac{1}{n_j} - c'(k_j) = 0 \text{ for } j = F, I \quad (4.2.7)$$

Equation (4.2.6) is a familiar Cournot first-order condition. It indicates that the firm should increase production until its per-unit profit from expanding output by another unit ($\theta R^j - G(Q) - c(k_j)$) just equals the losses ($q_j G'(Q)$) this expansion would impose because of the induced price reduction on the other units the firm was selling. Equation (4.2.7) indicates that each producer in a given sector raises the quality of its output until the marginal cost of further expansion ($c'(k_j)$) equals the marginal benefit of that expansion ($\frac{\theta}{n_j}$). Note that a producer incurring the additional cost of increasing quality captures only $1/n_j^{th}$ of the benefit since the firm produces only $1/n_j$ of sector j 's output.

We assume that in the short run, the number of producers in each sector is set at its exogenously specified level. However, we suppose that in the long run it is costless for a producer to move between the formal and informal sectors. This leads to the following result:

Theorem 4.2.1 *The laissez-faire benchmark: in the short run, the output from the sector with the larger number of producers will be of lower quality and will sell for a lower price; firms operating in this sector will produce less output and make smaller profits. In the long run, the number of producers operating in each sector will equalize and all firms will produce the same volume of output; quality, price and profits will equalize among producers in the two sectors.*

Proof. Equation (4.2.7) implies that the sector with the larger number of firms will offer lower-quality experience goods. Because prices adjust so that consumers are indifferent about the source of the good, output in the sector with the larger number of producers must sell at a lower price. This in turn ensures that every firm operating in the larger sector has lower profit and output, as the following argument shows. In the equilibrium, any firm offering quality k earns profit per unit of $\theta k - G(Q) - c(k)$. As this function is strictly concave in quality and peaks at k^* , the implicit solution to $\theta = c'(*)$, equation (4.2.7) implies that in

equilibrium, every firm will choose a quality $k_j < k^*$. Therefore, the profit per unit at each firm in a group rises if the common quality of every firm in that group increases. It follows that firms in the larger sector will have lower profit per unit than firms in the smaller sector. But equation (4.2.6) implies that any firm with a lower profit per unit produces less output and hence earns lower total profit.¹⁰ In the limiting case where each sector has the same number of firms, quality, price, output, and profit are the same at every firm regardless of its sector.

In long-run equilibrium, because switching between sectors is costless, producers will reallocate themselves until the profits earned in the formal market are equal to the profits earned in the informal sector. Moreover, producers in each sector are free to adjust their output until the *per-unit profit* of expanding output equals the losses that would be imposed on the inframarginal units sold by a firm. Consequently, profits at a firm in sector j will equal $q_j^2 G'(Q)$ for $j = F, I$. It follows that in the long-run equilibrium, output will equalize among firms in the two sectors, and so will profit per unit. This implies that quality must be equalized across sectors and so number of firms must be equalized across sectors. ■

4.3 Consequences of Two Policies

Suppose that the formal and informal sectors are in long-run equilibrium so that the number of firms operating in each sector is equal, and profits, output, and qualities equalize across sectors. Assume the government enacts a regulation binding only on firms in the formal sector. We will first consider the impact of a constant tax per-unit (τ) on firms operating in the formal sector. We have the following result:

Theorem 4.3.1 *If an equal number of firms are initially located in each sector, then in the short run, the imposition of a per-unit tax ($\tau > 0$) on the formal sector will leave quality unchanged in both sectors and will decrease aggregate output. Output, profit per firm, and profit per unit will decrease in the formal sector and increase in the informal sector.*

In the long run, producers will migrate from the formal to the informal sector, and those remaining in the formal sector will offer higher qualities and higher prices than firms in the informal sector. Output per firm expands in the formal sector and contracts in the informal sector. Profits equalize across sectors.

¹⁰In this model, it is assumed that output produced in the formal sector is marketed in the formal sector and output produced in the informal sector is marketed in the informal sector. This rules out firms producing in one sector and then disguising their products as originating in the other sector so as to earn higher prices and profits. It also rules out purchasing a good in the sector with the lower price and reselling it in the sector with the higher price.

Proof. When the regulations are imposed, each sector has an equal number of producers. Given the imposition of the tax, we can write the first-order conditions of the firms in the formal sector as:

$$\theta R^F - G(Q) - (c(k_F) + \tau) - q_F G'(Q) = 0 \quad (4.3.1)$$

$$\theta \frac{1}{n_F} - c'(k_F) = 0 \quad (4.3.2)$$

The first-order conditions of firms in the informal sector are unchanged. From this, one can immediately see that in the short run, the imposition of the tax has no impact on the quality choices of the firms. We can show that:

$$\frac{dq_F}{d\tau} = \frac{-(G'(Q) + q_F G''(Q)) \frac{dQ}{d\tau} - 1}{G'(Q)} < 0 \quad (4.3.3)$$

$$\frac{dq_I}{d\tau} = \frac{-(G'(Q) + q_I G''(Q)) \frac{dQ}{d\tau}}{G'(Q)} > 0 \quad (4.3.4)$$

Using the fact that $\frac{dQ}{d\tau} = n_F \frac{dq_F}{d\tau} + n_I \frac{dq_I}{d\tau}$, we get that:

$$\frac{dQ}{d\tau} = \frac{-n_F}{(1 + n_F + n_I)G'(Q) + (n_F q_F + n_I q_I)G''(Q)} < 0 \quad (4.3.5)$$

As (12) indicates, the tax must decrease aggregate production; every term in the denominator of (12) is strictly positive. Given this induced aggregate contraction, the tax must cause every firm in the informal sector to expand (as (11) indicates). As for firms in the formal sector, the tax must cause them to contract as (10) reflects (otherwise every firm would expand, contradicting the result established above that aggregate production contracts). Because both factors in the last term of equation (4.3.1) decrease, after-tax profit per unit in the formal market decreases as well. Thus, profits in the formal sector decrease. Profits must increase in the informal sector. This follows because the tax raises output per firm in the informal sector and—because consumer utility decreases—must result in a higher price and profit per unit.

In the long run, producers operating in the formal sector will move to the informal sector to take escape the tax. In other words, n_F will decrease and n_I will increase. From the first-order conditions, it is easy to see that the reputed quality will increase in the formal market and will decrease in the informal market. In particular, we have

$$\frac{dk_F}{dn_I} = \frac{c'(k_F)}{n_F c''(k_F)} > 0, \quad \frac{dk_I}{dn_I} = \frac{-c'(k_I)}{n_I c''(k_I)} < 0 \quad (4.3.6)$$

We can also calculate that

$$\frac{dq_F}{dn_I} = \frac{(\theta - c'(k_F)) \frac{dk_F}{dn_I} - (G'(Q) + q_F G''(Q)) \frac{dQ}{dn_I}}{G'(Q)} \quad (4.3.7)$$

$$\frac{dq_I}{dn_I} = \frac{(\theta - c'(k_I)) \frac{dk_I}{dn_I} - (G'(Q) + q_I G''(Q)) \frac{dQ}{dn_I}}{G'(Q)} \quad (4.3.8)$$

Aggregate output may either expand or contract in response to the reallocation of firms across the two sectors. However, in both cases, output per firm expands in the formal sector and contracts in the informal sector.¹¹ Firms will continue to move from the informal sector to the formal sector until profits are equalized. In long-run equilibrium, the formal sector will be smaller and will offer higher quality goods at higher price relative to the informal sector. ■

This result illustrates the surprising effects of the ability of firms to move in and out of the formal and informal sectors. Due to the additive separability of the inverse demand function, a tax affects only the quantity decisions of producers in the formal sector but the changes in volume do not induce any changes in the quality of the goods produced. In the short run, the tax also has no effect on quality in the informal sector since the number of firms in that sector does not change. However, the tax does still affect the profits of producers operating in the formal sector, which in the long run incentivizes firms operating in the formal sector to move to the informal sector. This has the effect of raising the reputed quality in the formal market and lowering the reputed quality in the informal market.¹²

We now analyze the case where the government decides to regulate the quality of the product in the formal market in the form of a minimum quality standard.¹³ We assume that the standard (\bar{k}) is binding on firms in the formal sector but is not set higher than the firm would choose if it were the only domestic producer, and hence its products could be readily identified by consumers.

That is, we assume $\bar{k} \leq k^*$, where k^* solves $\theta = c'(\cdot)$. Note that k^* is also the optimal quality chosen by a social planner to maximize welfare. We establish the following:

Theorem 4.3.2 *In the short run, the imposition of a minimum quality standard on the formal sector raises*

¹¹To see this, suppose that aggregate output increases. Since $\theta \geq c'(k_I) \geq 0$, $\frac{dk_I}{dn_I} < 0$, and $G(Q)$ is strictly increasing and convex, it is clear from equation (15) that output contracts in the informal sector. But then if aggregate output increases, it must be that output increases in the formal sector. Similarly, if aggregate output decreases, one can show from equation (14) that output increases in the formal sector, which implies that output decreases in the informal sector.

¹²The effect is even more surprising if the government subsidizes formal sector production ($\tau < 0$). Such a subsidy would actually degrade the quality of goods *within* the ambit of regulation and would effectively raise the quality of goods *outside* the ambit of regulation.

¹³One form this regulation could take is to require the sellers in the formal market to give their customers a receipt. If the quality of the good was less than that imposed by the standard, the customer could use the receipt to file a legal claim against the seller. If the cost of these suits to the sellers were sufficiently high, sellers would refuse to deal with producers of goods with quality less than that imposed by the government standard. Thus, requiring that sellers in the formal sector furnish receipts is an indirect way of imposing a minimum quality standard on producers in the formal sector.

the output and profits of firms in the formal sector while lowering the output and profits of unregulated firms in the informal sector. Overall, aggregate output expands. Quality rises in the formal sector and remains unchanged elsewhere.

In the long run, firms move from the informal sector to the formal sector, and profits, profit per unit, output, and quality equalize across sectors.

Proof. The imposition of the standard must strictly increase aggregate production of the experience good. For, suppose the contrary. Suppose aggregate quantity falls or remains constant. Then the utility that consumers get from the experience good must weakly decrease. In the informal sector, firms would maintain quality, as equation (4.2.7) still holds. So if their goods provide weakly less net utility, the prices of their goods ($P^j = \theta k_j - U(Q)$) must weakly increase. Because the per-unit profit ($P^j - c(k_j)$) would then weakly increase, equation (4.2.6) implies that output at each unregulated firm must weakly increase. As for the *regulated* firms in the formal sector, their per-unit profit must *strictly* increase, because the standard raised quality and, by assumption, was not excessive ($\bar{k} < k^*$). Equation (4.2.6) then implies that output at each regulated firm strictly increases. But then we have a contradiction: aggregate output cannot weakly decrease as we hypothesized, as this implies that the sum of the individual firm outputs would strictly increase.

So the imposition of a minimum quality standard in one market must cause aggregate output of the experience good to strictly increase and hence must cause the net utility of every consumer of the good to increase. As the quality of the unregulated firms in the informal sector does not change, their prices, profit per unit, output, and total profits must fall. Because aggregate output expands despite the contraction at every unregulated firm, output at every *regulated* firm must increase. But, as equation (4.2.6) implies, regulated firms would expand output only if their profit per unit also increased. Hence, their total profits would also increase. Because profit per unit increases at each regulated firm, its price per unit must increase by more than enough to offset the increased cost per unit of producing the higher quality mandated by the minimum quality standard.

In the long run, firms will abandon the informal sector and join the formal sector. This will raise the quality of the product in the informal market, while the quality in the formal market remains fixed at the level imposed by the minimum quality standard. Note that equation (4.2.6) implies that if profit per unit is higher in one sector, then output is higher in that sector, which implies profit is higher. Thus, in long-run equilibrium, profit per unit and output must be equalized across sectors. Because profit per unit ($\theta k_j - U(Q) - c(k_j)$, for $j = F, I$) is equalized across sectors, this implies that quality must also be equalized across sectors. ■

This result is quite interesting. It demonstrates that given the lack of firm traceability with regard to quality, the government can indirectly, yet effectively, regulate the *entire* market for a good through regulations on only the formal sector. Intuitively, the minimum quality standard actually creates a comparative advantage for the formal sector, which leads firms to move out of the informal sector. As firms leave the informal sector, the collective reputation problem is ameliorated, which endogenously raises the quality of goods in this sector. This process continues until the quality of good in the informal market is exactly equal to the quality specified by the standard.

By setting the standard at k^* , the government can in fact achieve the socially optimal quality economy-wide, in both the formal and informal sectors. Even though the informal sector is outside the regulatory reach, spillover effects from the regulation on the formal sector and market forces drive quality in this sector to the socially optimal quality.¹⁴

4.4 Conclusion

In his V.V. Giri Memorial Lecture, Ravi Kanbur (2009) discusses the “informality discourse” in the development literature and notes that the literature would benefit from a more precise specification of the regulations imposed in the formal sector as well as from a recognition that such regulations on firms in the formal sector may induce them to “adjust activity to move out of the ambit of the regulation.”

We adopt his suggestion about specifying regulations more precisely. This paper assumes that consumers can purchase a product on either the formal or informal market and can assess the quality of that product only from the collective reputation of the sector where it is produced and sold. In this circumstance, firms will free ride on quality efforts of others in that sector, and as a result, quality provision in each sector will be suboptimal. We show that regulations imposed on only formal sector firms can generate a quality gap between formal market and informal market goods of the class considered. In particular, if a per-unit tax is imposed, some producers will eventually migrate to the informal sector. The reduced free riding among formal sector producers will result in higher quality goods and elevated prices as compared to the informal sector.

The paper additionally demonstrates, however, that regulations on formal sector firms need not always lead to a quality gap in which goods produced in the formal sector are of higher quality. If, for instance, regulators implement a minimum quality standard on producers in the formal sector, firms will be *attracted* to the regulated sector. The possibility that a regulation may *benefit* firms in the regulated sector has

¹⁴The policy is welfare-improving overall in the long run since it increases long run aggregate output. By the discussion above, in the long run the policy is equivalent to a minimum quality standard which is binding on *both* formal sector and informal sector firms. It is easy to show that such a policy increases aggregate output. For details see McQuade et al. (2012).

apparently received little attention.¹⁵ Under this policy, there is less free riding in the informal sector because fewer firms remain in it *and* less free riding in the formal sector as a result of the quality standard. Therefore, a minimum quality standard imposed on firms in the formal sector will increase quality in both sectors to the level specified by the standard and will, in fact, *eliminate* any existing quality gap.

¹⁵The idea of a government intervention in one area *benefiting* agents in another area despite the absence of any intervention in that area has other applications. For example, if a fixed number of motorists commute from the same origin to the same destination on two congestible roads, then expanding one road so that commute time for any given number of cars is reduced will attract motorists from the other road and will therefore reduce the commute time on that other road, even though the government has made no improvements to that road. One can think of other examples where everyone is helped by regulation even those who are not directly subject to it. In the case of the highway expansion, it is quite plausible that commute time will improve on both roads to exactly the same extent. In the case of a minimum quality standard in the formal market, quality is likely to rise in the regulated market. Whether it rises to the *same* extent in the informal market as our model predicts depends on factors from which we have abstracted. Although a standard might in reality affect quality in the informal market less than in the formal market, the movement of quality in the same direction in the two markets might make identifying empirically the full effects of the regulation a challenge.

Bibliography

- ADRIAN, T., AND J. ROSENBERG (2008): “Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk,” Journal of Finance, 63, 2997–3030.
- AKERLOF, G. A. (1970): “The Market for “Lemons”: Quality Uncertainty and the Market Mechanism,” Quarterly Journal of Economics, 84, 488–500.
- ALBRECHT, J., P. GAUTIER, AND S. VROMAN (2010): “Directed Search in the Housing Market,” Georgetown University Working Paper.
- ALIZADEH, S., M. BRANDT, AND F. DIEBOLD (2002): “Range-Based Estimation of Stochastic Volatility Models,” Journal of Finance, 57, 1047–1091.
- ANENBERG, E., AND P. BAYER (2012): “Endogenous Sources of Cycling in Housing Markets: The Joint Buyer-Seller Problem,” Duke University Working Paper.
- ANENBERG, E., AND E. KUNG (2012): “Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures,” Duke University Working Paper.
- ANG, A., R. HODRICK, Y. XING, AND X. ZHANG (2006): “The Cross-Section of Volatility and Expected Returns,” Journal of Finance, 61, 259–299.
- AVRAMOV, D., T. CHORDIA, G. JOSTOVA, AND A. PHILOPOV (2007): “Momentum and Credit Rating,” Journal of Finance, 62, 2503–2520.
- BAJARI, P., S. CHU, AND M. PARK (2010): “An Empirical Model of Subprime Mortgage Default from 2000 to 2007,” University of Minnesota Working Paper.
- BANERJI, A., AND S. JAIN (2007): “Quality Dualism,” Journal of Development Economics, 84, 234–250.

- BANSAL, R., D. KIKU, I. SHALIASTOVICH, AND A. YARON (2012): “Volatility, the Macro-economy and Asset Prices,” working paper, Duke University and University of Pennsylvania.
- BANSAL, R., AND A. YARON (2004): “Risks for the Long Run,” Journal of Finance, 59, 1481–1509.
- BASU, S. (1977): “The Investment Performance of Common Stocks in Relation to Their Price to Earnings Ratios: A Test of the Efficient Markets Hypothesis,” Journal of Finance, 32, 663–682.
- (1983): “The Relationship between Earnings Yield, Market Value, and Return for NYSE Common Stocks: Further Evidence,” Journal of Financial Economics, 12, 129–156.
- BHUTTA, N., J. JOKKO, AND H. SHAN (2011): “Consumer Ruthlessness and Mortgage Default During the 2007-2009 Housing Bust,” Federal Reserve Board Working Paper.
- BIAIS, B., T. MARIOTTI, J. ROCHET, AND S. VILLENEUVE (2010): “Large Risks, Limited Liability, and Dynamic Moral Hazard,” Econometrica, 78, 73–118.
- BLACK, F., AND J. COX (1976): “Valuing Corporate Securities: Some Effects of Bond Indenture Provisions,” Journal of Finance, 31, 351–367.
- BLACK, F., AND M. SCHOLES (1973): “The Pricing of Options and Corporate Liabilities,” Journal of Political Economy, 81, 637–654.
- CALOMIRIS, C., S. LONGHOFFER, AND W. MILES (2008): “The Foreclosure-House Price Nexus: Lessons from the 2007-2008 Housing Turmoil,” Working Paper 12918, NBER.
- CALVET, L., AND A. FISHER (2001): “Forecasting Multifractal Volatility,” Journal of Econometrics, 105, 27–58.
- (2004): “How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes,” Journal of Financial Econometrics, 2, 49–83.
- CAMPBELL, J. (1993): “Intertemporal Asset Pricing Without Consumption Data,” American Economic Review, 83, 487–512.
- (1996): “Understanding Risk and Return,” Journal of Political Economy, 104, 298–345.
- CAMPBELL, J., AND J. COCCO (2011): “A Model of Mortgage Default,” Working Paper 17516, NBER.
- CAMPBELL, J., S. GIGLIO, AND P. PATHAK (2011): “Forced Sales and House Prices,” American Economic Review, 101(5), 2108–31.

- CAMPBELL, J., S. GIGLIO, C. POLK, AND R. TURLEY (2012): “An Intertemporal CAPM with Stochastic Volatility,” working paper, Harvard University.
- CAMPBELL, J., J. HILSCHER, AND J. SZILAGYI (2008): “In Search of Distress Risk,” Journal of Finance, 63, 2899–2939.
- CAMPBELL, J., AND G. TAKSLER (2003): “Equity Volatility and Corporate Bond Yields,” Journal of Finance, 58, 2321–2350.
- CARLSON, M., A. FISHER, AND R. GIAMMARINO (2004): “Corporate Investment and Asset Pricing Dynamics: Implications for the Cross-Section of Returns,” Journal of Finance, 59, 2577–2603.
- CASTELLS, M., AND A. PORTES (1989): “World Underneath: The Origins, Dynamics and Effects of the Informal Economy,” in The Informal Economy: Studies in Advanced and Less Developed Economies, ed. by A. Portes, M. Castells, and L. A. Benton, pp. 11–37. JHU Press, Baltimore, MD.
- CHACKO, G., AND L. VICEIRA (2005): “Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets,” Review of Financial Studies, 18, 1369–1402.
- CHATTERJEE, S., AND B. EYIGUNGOR (2011): “A Quantitative Analysis of the U.S. Housing and Mortgage Markets and the Foreclosure Crisis,” Working paper, Federal Reserve Bank of Philadelphia.
- CHEN, H. (2010): “Macroeconomic Conditions and the Puzzle of Credit Spreads and Capital Structure,” Journal of Finance, 65, 2171–2212.
- CHEN, N., AND S. KOU (2009): “Credit Spreads, Optimal Capital Structure, and Implied Volatility with Endogenous Default and Jump Risk,” Mathematical Finance, 19, 343–378.
- CHERNOV, M., R. GALLANT, E. GHYSELS, AND G. TAUCHEN (2003): “Alternative Models for Stock Price Dynamics,” Journal of Econometrics, 116, 225–257.
- CHETTY, R., AND A. SZEIDL (2007): “Consumption Commitments and Risk Preferences,” Quarterly Journal of Economics, 122(2), 831–877.
- CHIANG, S.-C., AND R. T. MASSON (1988): “Domestic Industrial Structure and Export Quality,” International Economic Review, 29, 261–270.
- CLAURETIE, T., AND N. DANESHVARY (2009): “Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time,” Real Estate Economics, 37(1), 43–67.
- COCHRANE, J. (2011): “Presidential Address: Discount Rates,” Journal of Finance, 66, 1047–1108.

- COLLIN-DUFRESNE, P., AND R. GOLDSTEIN (2001): “Do Credit Spreads Reflect Stationary Leverage Ratios?,” Journal of Finance, 56, 1929–1957.
- COLLIN-DUFRESNE, P., R. GOLDSTEIN, AND J. MARTIN (2001): “The Determinants of Credit Spread Changes,” Journal of Finance, 56, 2177–2207.
- COOPER, I. (2006): “Asset Pricing Implications of Nonconvex Adjustment Costs and Irreversibility of Investment,” Journal of Finance, 61, 139–170.
- COVAL, J., AND T. SHUMWAY (2001): “Expected Option Returns,” Journal of Finance, 56, 983–1009.
- DENG, Y., J. M. QUIGLEY, AND R. V. ORDER (2000): “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options,” Econometrica, 68(2), 275–307.
- DESSY, S., AND S. PALLAGE (2003): “Taxes, Inequality and the Size of the Informal Sector,” Journal of Development Economics, 70, 225–233.
- DIAZ, A., AND B. JEREZ (2009): “House Prices, Sales and Time on the Market: A Search-Theoretic Framework,” Working Paper 09-25, Universidad Carlos II de Madrid.
- DIXIT, A. K. (2004): Lawlessness and Economics: Alternative Models of Governance. Princeton University Press, Princeton.
- DONNENFELD, S., AND W. MAYER (1987): “The Quality of Export Products and Optimal Trade Policy,” International Economic Review, 28, 159–174.
- DUFFEE, G. (1998): “The Relation between Treasury Yields and Corporate Bond Yield Spreads,” Journal of Finance, 53, 2225–2241.
- (1999): “Estimating the Price of Default Risk,” Review of Financial Studies, 12, 197–226.
- DUFFIE, D., N. GARLEANU, AND L. PEDERSEN (2007a): “Valuation in Over-the-Counter Markets,” Review of Financial Studies, 20, 1865–1900.
- (2007b): “Valuation in Over-the-Counter Markets,” Review of Financial Studies, 20, 1865–1900.
- DUFFIE, D., AND D. LANDO (2001): “Term Structures of Credit Spreads with Incomplete Accounting Information,” Econometrica, 69, 633–664.
- DUFFIE, D., AND K. SINGLETON (1999): “Modeling Term Structures of Defaultable Bonds,” Review of Financial Studies, 12, 687–720.

- ELTON, E. J., M. J. GRUBER, D. AGRAWAL, AND C. MANN (2001): “Explaining the Rate Spread on Corporate Bonds,” Journal of Finance, 56, 247–277.
- ELUL, R., N. S. SOULELES, S. CHOMSISENGPHET, D. GLENNON, AND R. HUNT (2010): “What ‘Triggers’ Mortgage Default?,” Working Paper 10-13, Federal Reserve Bank of Philadelphia.
- ENGLE, R. (1982): “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,” Econometrica, 50, 987–1007.
- ENGLE, R., AND J. ROSENBERG (2000): “Testing the Volatility Term Structure Using Option Hedging Criteria,” Journal of Derivatives, 8, 10–28.
- ERAKER, B. (2004): “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices,” Journal of Finance, 59, 1367–1403.
- FALVEY, R. E. (1989): “Quality Reputations and Commercial Policy,” International Economic Review, 30, 607–622.
- FAMA, E., AND K. FRENCH (1993): “Common Risk Factors in the Returns on stocks and Bonds,” Journal of Financial Economics, 33, 3–56.
- (1996): “Multifactor Explanations of Asset Pricing Anomalies,” Journal of Finance, 51, 55–84.
- FERREIRA, F., J. GYOURKO, AND J. TRACY (2010): “Housing Busts and Household Mobility,” Journal of Urban Economics, 68(1), 34–45.
- FOOTE, C. L., K. GERARDI, L. GOTTE, AND P. S. WILLEN (2009): “Reducing Foreclosures,” Discussion Paper 09-2, Federal Reserve Bank of Boston.
- FOOTE, C. L., K. GERARDI, AND P. S. WILLEN (2008): “Negative Equity and Foreclosure: Theory and Evidence,” Journal of Urban Economics, 64(2), 234–245.
- FORTIN, B., N. MARCEAU, AND L. SAVARD (1997): “Taxation, Wage Controls and the Informal Sector,” Journal of Public Economics, 66, 293–312.
- FOUQUE, J., G. PAPANICOLAOU, R. SIRCAR, AND K. SÄYLÄ (2003a): “Multiscale Stochastic Volatility Asymptotics,” Multiscale Modeling & Simulation, 2, 22–42.
- (2003b): “Short Time-Scale in S&P 500 Volatility,” Journal of Computational Finance, 6, 1–24.
- (2003c): “Singular Perturbations in Option Pricing,” SIAM Journal on Applied Mathematics, 63, 1648–1665.

- (2004): “Stochastic Volatility Corrections for Interest Rate Derivatives,” Mathematical Finance, 14, 173–200.
- (2006): “Stochastic Volatility Effects on Defaultable Bonds,” Applied Mathematical Finance, 13, 215–244.
- (2011): Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives. Cambridge University Press, Cambridge, England.
- GARLAPPI, L., AND H. YAN (2011): “Financial Distress and the Cross-Section of Equity Returns,” Journal of Finance, 66, 789–822.
- GENESOVE, D., AND L. HAN (2012): “Search and Matching in the Housing Market,” Journal of Urban Economics, 72(1), 31–45.
- GENESOVE, D., AND C. MAYER (2001): “Loss Aversion and Seller Behavior: Evidence From the Housing Market,” Quarterly Journal of Economics, 116(4), 1233–1260.
- GERARDI, K., E. ROSENBLATT, P. S. WILLEN, AND V. W. YAO (2012): “Foreclosure Externalities: Some New Evidence,” Federal Reserve Bank of Boston Working Paper.
- GOLDSTEIN, R., N. JU, AND H. LELAND (2001): “An EBIT-Based Model of Dynamic Capital Structure,” Journal of Business, 74, 483–512.
- GOMES, J., L. KOGAN, AND L. ZHANG (2003): “Equilibrium Cross Section of Returns,” Journal of Political Economy, 111, 693–732.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2009): “Moral and Social Constraints to Strategic Default on Mortgages,” Working Paper 15145, NBER.
- HACKBARTH, D., J. MIAO, AND E. MORELLEC (2006): “Capital Structure, Credit Risk, and Macroeconomic Conditions,” Journal of Financial Economics, 82, 519–550.
- HELLWIG, M. (1981): “Bankruptcy, Limited Liability, and the Modigliani-Miller Theorem,” American Economic Review, 71, 155–170.
- HESTON, S. (1993): “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options,” Review of Financial Studies, 6, 327–343.
- HILBERINK, B., AND C. ROGERS (2002): “Optimal Capital Structure and Endogenous Default,” Finance and Stochastics, 6, 237–263.

- HUANG, J., AND M. HUANG (2003): “How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk?,” working paper, Stanford University.
- IACOVIELLO, M. (2005): “House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle,” American Economic Review, 95, 739–764.
- IHRIG, J., AND K. S. MOE (2004): “Lurking in the Shadows: The Informal Sector and Government Policy,” Journal of Development Economics, 73, 541–557.
- INNES, R. (1990): “Limited Liability and Incentive Contracting with Ex-Ante Action Choices,” Journal of Economic Theory, 52, 45–67.
- JEGADEESH, N., AND S. TITMAN (1993): “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” Journal of Finance, 48, 65–91.
- JONES, C. (2003): “The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets,” Journal of Econometrics, 116, 181–224.
- KANBUR, R. (2009): “Conceptualising Informality: Regulation and Enforcement,” Indian Journal of Labour Economics, 52, 33–42.
- KIYOTAKI, N., AND J. MOORE (1997): “Credit Cycles,” Journal of Political Economy, 105(2), 211–248.
- KLEIN, B., AND K. B. LEFFLER (1981a): “The Role of Market Forces in Assuring Contractual Performance,” Journal of Political Economy, 89, 615–641.
- (1981b): “The Role of Market Forces in Assuring Contractual Performance,” Journal of Political Economy, 89, 615–641.
- KORTEWEG, A., AND M. SORENSEN (2011): “Estimating Loan-to-Value and Foreclosure Behavior,” Columbia University Working Paper.
- KRAINER, J. (2001): “A Theory of Liquidity in Residential Real Estate Markets,” Journal of Urban Economics, 49, 32–52.
- LAMONT, O., AND J. C. STEIN (1999): “Leverage and House-Price Dynamics in U.S. Cities,” RAND Journal of Economics, 30(3), 498–514.
- LANDVOIGT, T., M. PIAZZESI, AND M. SCHNEIDER (2012): “The Housing Market(s) of San Diego,” Stanford University Working Paper.

- LAUX, C. (2001): “Limited-Liability and Incentive Contracting with Multiple Projects,” RAND Journal of Economics, 32, 514–526.
- LEE, R. (2001): “Implied and Local Volatilities under Stochastic Volatility,” International Journal of Theoretical and Applied Finance, 1, 45–89.
- LELAND, H. (1994a): “Bond Prices, Yield Spreads, and Optimal Capital Structure with Default Risk,” working paper 240, Haas School of Business, U.C. Berkeley.
- (1994b): “Corporate Debt Value, Bond Covenants, and Optimal Capital Structure,” Journal of Finance, 49, 157–196.
- (1998): “Agency Costs, Risk Management, and Capital Structure,” Journal of Finance, 53, 1213–1243.
- LELAND, H. (2004): “Predictions of Default Probabilities in Structural Models,” Journal of Investment Management, 2, 1–16.
- LELAND, H., AND K. TOFT (1996): “Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads,” Journal of Finance, 51, 987–1019.
- LIU, J. (2007): “Portfolio Selection in Stochastic Environments,” Review of Financial Studies, 20, 1–39.
- LIVINGSTONE, I. (1991): “A Reassessment of Kenya’s Rural and Urban Informal Sector,” World Development, 19, 651–670.
- LONGSTAFF, F., AND E. SCHWARTZ. (1995): “A Simple Approach to Valuing Risky Fixed and Floating Rate Debt,” Journal of Finance, 50, 789–819.
- MALLOY, R., AND H. SHAN (2011): “The Post-Foreclosure Experience of U.S. Households,” Federal Reserve Board Working Paper.
- MARJIT, S., S. GHOSH, AND A. BISWAS (2007): “Informality, corruption and trade reform,” European Journal of Political Economy, 23, 777–789.
- MARJIT, S., AND S. KAR (forthcoming): The Outsiders: Economic Reform and Informal Labour in a Developing Economy. Oxford University Press, Oxford.
- MAYER, C., K. PENCE, AND S. M. SUNDERLUND (2009): “The Rise in Mortgage Default,” Journal of Economic Perspectives, 23(1), 27–50.

- MAZUMDAR, D. (1983): “Segmented Labour Markets in LDCs,” American Economic Review, 73, 254–259.
- MERTON, R. (1973): “An Intertemporal Capital Asset Pricing Model,” Econometrica, 41, 867–887.
- (1974): “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates,” Journal of Finance, 29, 449–470.
- MIAN, A., AND A. SUFI (2009): “The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis,” Quarterly Journal of Economics, 124(4), 1449–1496.
- (2011): “House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis,” American Economic Review, 101(5), 2132–2156.
- MIAN, A., A. SUFI, AND F. TREBBI (2011): “Foreclosures, House Prices, and the Real Economy,” Working Paper 16685, NBER.
- MYINT, H. (1985): “Organizational Dualism and Economic Development,” Asian Development Review, 3, 24–42.
- NELSON, P. (1970): “Information and Consumer Behavior,” Journal of Political Economy, 78, 311–329.
- NGAI, L. R., AND S. TENREYRO (2010): “Hot and Cold Seasons in the Housing Market,” Discussion Paper 922, Centre for Economic Performance.
- NOVY-MARX, R. (2009): “Hot and Cold Markets,” Real Estate Economics, 37(1), 1–22.
- ORGANISATION, I. L. (1972): Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya. ILO, Geneva.
- ORTALO-MAGNE, F., AND S. RADY (2006): “Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints,” Review of Economic Studies, 73, 459–485.
- P., C., AND L. WU (2009): “Variance Risk Premiums,” Review of Financial Studies, 22, 1311–1341.
- PAN, J. (2002): “The jump-risk premia implicit in options: evidence from an integrated time-series study,” Journal of Financial Economics, 63, 3–50.
- PAPANIKOLAOU, D. (2011): “Investment Shocks and Asset Prices,” Journal of Political Economy, 119, 639–685.
- PIAZZESI, M., AND M. SCHNEIDER (2009): “Momentum Traders in the Housing Market: Survey Evidence and a Search Model,” American Economic Review, 99(2), 406–411.

- RAUCH, J. E. (1991): “Modelling the Informal Sector Formally,” Journal of Development Economics, 35, 33–47.
- RESEARCH, Z. R. E. (2010): “Price Differences Between Foreclosures and Non-Foreclosures,” .
- SAINAI, T., AND N. S. SOULELES (2009): “Can Owning a Home Hedge the Risk of Moving?,” University of Pennsylvania Working Paper.
- SCHAEFER, S., AND I. STREBULAIEV (2008): “Structural Models of Credit Risk are Useful: Evidence from Hedge Ratios on Corporate Bonds,” Journal of Financial Economics, 90, 1–19.
- SHAPIRO, C. (1983): “Premiums for High Quality Products as Returns to Reputations,” Quarterly Journal of Economics, 98, 659–680.
- SHIMER, R., AND I. WERNING (2007): “Reservation Wages and Unemployment Insurance,” Quarterly Journal of Economics, 122(3), 1145–1185.
- SPRINGER, T. M. (1996): “Single-Family Housing Transactions: Seller Motivations, Price, and Marketing Time,” Journal of Real Estate Finance and Economics, 13, 237–254.
- STEIN, J. (1995): “Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects,” Quarterly Journal of Economics, 110(2), 379–406.
- T., B. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” Journal of Econometrics, 31, 307–327.
- T., B., M. GIBSON, AND H. ZHOU (2011): “Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities,” Journal of Econometrics, 160, 235–245.
- T., B., AND H. ZHOU (2002): “Estimating Stochastic Volatility Diffusions Using Conditional Moments of Integrated Volatility,” Journal of Econometrics, 109, 33–65.
- THEOLOGIDES, S. (2010): “Servicing REO Properties: The Servicer’s Role and Incentives,” in REO & Vacant Properties: Strategies for Neighborhood Stabilization.
- TIROLE, J. (1996): “A Theory of Collective Reputation (with Applications to the Persistence of Corruption and to Firm Quality),” Review of Economic Studies, 63, 1–22.
- VASSALOU, M., AND Y. XING (2004): “Default Risk in Equity Returns,” Journal of Finance, 59, 831–868.
- WHEATON, W. (1990): “Vacancy, Search, and Prices in a Housing Market Matching Model,” Journal of Political Economy, 98(6), 1270–1292.

- WILLIAMS, J. (1995): “Pricing Real Assets With Costly Search,” Review of Financial Studies, 8(1), 55–90.
- WINFREE, J., AND J. MCCLUSKEY (2005): “Collective Reputation and Quality,” American Journal of Agricultural Economics, 87, 206–213.
- ZHANG, B., H. ZOU, AND H. ZHU (2009): “Explaining Credit Default Swap Spreads with the Equity Volatility and Jump Risks of Individual Firms,” Review of Financial Studies, 22, 5099–5131.
- ZHANG, L. (2005): “The Value Premium,” Journal of Finance, 60, 67–103.
- ZHOU, C. (2001): “The Term Structure of Credit Spreads with Jump Risk,” Journal of Banking and Finance, 25, 2015–2040.

Appendix A

Supplement to Chapter 1

A.1 Derivations and Proofs

A.1.1 Proof of Theorem 1.3.1

I prove Theorem 1.3.1 in a sequence of lemmas. Let

$$DIV(x, y) = (1 - \phi)(x - C) + \tilde{d}(x, y) - p \quad (\text{A.1.1})$$

denote the flow dividend which accrues to the equityholders and set:

$$E^*(x, y) = \sup_{\tau \in T} \mathbb{E}_t^* \left[\int_t^\tau e^{-r(s-t)} DIV(X_s, Y_s) ds \right]. \quad (\text{A.1.2})$$

The goal is to show that the solution $E(x, y)$ to equations (1.3.22a)-(1.3.22h) satisfies $E(x, y) = E^*(x, y)$ for all $x, y > 0$. Let $\mathcal{C} = \{(x, y) : x > x_B(y)\}$ denote the continuation set and $\mathcal{D} = \{(x, y) : x \leq x_B(y)\}$ the stopping set. Finally, define the stopping time:

$$\tau^B = \min \{t \geq 0 : (X_t, Y_t)\} \in \mathcal{D}. \quad (\text{A.1.3})$$

The first result is as follows:

Lemma A.1.1 *The flow dividend to equityholders must be nonpositive in the stopping region; that is, $DIV(x, y) \leq 0$ for all $(x, y) \in \mathcal{D}$.*

Proof. The proof is very simple. Recalling equation (1.3.22b), we must have:

$$DIV(x, y) + \mathcal{L}_{X,Y} E \leq rE$$

for all $(x, y) \in \mathcal{D}$. But equation (1.3.22c) says that $E(x, y) = 0$ for all $(x, y) \in \mathcal{D}$. Substituting this into the equation above immediately gives the desired result. ■

This result is intuitive. Since equityholders receive nothing in the event of default, it can never be optimal for them to default when they are still receiving positive dividends. Now define:

$$Z_t = e^{-rt} E(X_t, Y_t) + \int_0^t e^{-rs} DIV(X_s, Y_s) ds. \quad (\text{A.1.4})$$

Lemma A.1.2 *The process Z_t satisfies:*

$$\int_0^t e^{-rs} DIV(X_s, Y_s) ds \leq Z_t \leq Z_0 + M_t \quad (\text{A.1.5})$$

where M_t is a continuous local martingale. The stopped process $Z_{t \wedge \tau^B}$ satisfies:

$$Z_{t \wedge \tau^B} = Z_0 + M_{t \wedge \tau^B}. \quad (\text{A.1.6})$$

Proof. The first inequality in equation (A.1.5) follows directly from the fact that $E(x, y) \geq 0$ for all x, y . For the second inequality, note that the Ito-Doebelin formula applies to functions whose second derivatives are discontinuous on measure zero sets as long as the first-derivatives are everywhere continuous, which itself is a consequence of smooth-pasting. Therefore, under the risk-neutral measure \mathbb{P}^* :

$$dZ_t = e^{-rt} \left[\{-rZ_t + \mathcal{L}_{X,Y} Z_t + DIV_t\} dt + \sqrt{Y_t} X_t \frac{\partial E}{\partial x} dW_t^{(1)*} + \nu_Y \sqrt{Y_t} \frac{\partial E}{\partial y} dW_t^{(2)*} \right]. \quad (\text{A.1.7})$$

By equation (1.3.22a), this is:

$$dZ_t = e^{-rt} DIV_t \mathbb{I}[X_t \leq x_B(Y_t)] dt + e^{-rt} \left[\sqrt{Y_t} X_t \frac{\partial E}{\partial x} dW_t^{(1)*} + \nu_Y \sqrt{Y_t} \frac{\partial E}{\partial y} dW_t^{(2)*} \right]. \quad (\text{A.1.8})$$

However, by the previous lemma I know that $DIV_t \leq 0$ for all $X_t \leq x_B(Y_t)$, which implies that:

$$Z_t \leq Z_0 + M_t \quad (\text{A.1.9})$$

where:

$$M_t = \int_0^t e^{-rs} \sqrt{Y_s} X_s \frac{\partial E}{\partial x} dW_s^{(1)*} + \int_0^t e^{-rs} \nu_Y \sqrt{Y_s} \frac{\partial E}{\partial y} dW_s^{(2)*} \quad (\text{A.1.10})$$

is a local continuous martingale by property of the Ito integral.

For the stopped process $Z_{t \wedge \tau^B}$, it is the case that $X_t > x_B(Y_t)$ for all $t < \tau^B$, which means that the indicator function in the drift term is equal to zero. Therefore:

$$Z_{t \wedge \tau^B} = Z_0 + M_{t \wedge \tau^B} \quad (\text{A.1.11})$$

as desired. ■

To complete the proof, I finally show:

Lemma A.1.3 *The solution $E(x, y)$ to equations (1.3.22a)-(1.3.22h) satisfies $E(x, y) = E^*(x, y)$*

Proof. Let $X_0 = x$ and $Y_0 = y$. Then for every stopping time τ and $n \in \mathbb{N}$:

$$\int_0^{\tau \wedge n} e^{-rs} DIV(X_s, Y_s) ds \leq E(x, y) + M_{\tau \wedge n}. \quad (\text{A.1.12})$$

By the optional sampling theorem, $\mathbb{E}_0^*[M_{\tau \wedge n}] = 0$, so that:

$$\mathbb{E}_0^* \left[\int_0^{\tau \wedge n} e^{-rs} DIV(X_s, Y_s) ds \right] \leq E(x, y). \quad (\text{A.1.13})$$

Taking limits as $n \rightarrow \infty$ and applying Fatou's lemma yields:

$$\mathbb{E}_0^* \left[\int_0^{\tau} e^{-rs} DIV(X_s, Y_s) ds \right] \leq E(x, y). \quad (\text{A.1.14})$$

Taking the supremum over all stopping times gives $E^*(x, y) \leq E(x, y)$.

For the other direction, by considering the stopped process $Z_{t \wedge \tau^B}$ and once again using the optional sampling theorem, I have that:

$$\mathbb{E}_0^* \left[e^{-r(\tau^B \wedge n)} E(X_{\tau^B \wedge n}, Y_{\tau^B \wedge n}) + \int_0^{\tau^B \wedge n} e^{-rs} DIV(X_s, Y_s) ds \right] = E(x, y). \quad (\text{A.1.15})$$

Taking limits as $n \rightarrow \infty$ and noting that $e^{-r\tau^B} E(X_{\tau^B}, Y_{\tau^B}) = 0$ gives:

$$E(x, y) = \mathbb{E}_0^* \left[\int_0^{\tau^B} e^{-rs} DIV(X_s, Y_s) ds \right]. \quad (\text{A.1.16})$$

This show that τ^B is optimal and that $E(x, y) = E^*(x, y)$ for all $x, y > 0$ as desired. ■

A.1.2 Derivation of Equation (1.4.13)

The problem for the principal order equity term is given by:

$$(1 - \phi)(x - C) + \tilde{d}_0^y(x) - p + \mathcal{L}_r^y E_0^y = 0 \quad \text{for } x > x_{B,0}^y \quad (\text{A.1.17a})$$

$$E_0^y(x_{B,0}^y) = 0 \quad (\text{A.1.17b})$$

$$\lim_{x \rightarrow \infty} E_0^y(x) = U(x) - \frac{C + mP}{r + m} + \frac{\phi C}{r} \quad (\text{A.1.17c})$$

$$\frac{\partial E_0^y}{\partial x} \Big|_{x=x_{B,0}^y} = 0 \quad (\text{A.1.17d})$$

To solve this problem I will introduce V_0^y as the solution to:

$$(1 - \phi)x + \phi C + \mathcal{L}_r^y V_0^y = 0 \quad \text{for } x > x_{B,0}^y \quad (\text{A.1.18a})$$

$$V_0^y(x_{B,0}^y) = (1 - \xi)U(x_{B,0}^y) \quad (\text{A.1.18b})$$

$$\lim_{x \rightarrow \infty} V_0^y(x) = U(x) + \frac{\phi C}{r} \quad (\text{A.1.18c})$$

Note that V_0^y is value of debt plus equity in the case of constant volatility. It is not equal to total firm value since the government is a residual claimant on a portion of the firm's cash flows. I now prove the following lemma:

Lemma A.1.4 *The principal equity value term $E_0^y(x) = V_0^y(x) - \tilde{d}_0^y(x)/m$.*

Proof. Recalling that $P = p/m$ and $C = c/m$, it is straightforward to check that:

$$V_0^y(x_{B,0}^y) - \tilde{d}_0^y(x_{B,0}^y)/m = 0 \quad (\text{A.1.19})$$

$$\lim_{x \rightarrow \infty} V_0^y(x) - \tilde{d}_0^y(x)/m = U(x) - \frac{C + mP}{r + m} + \frac{\phi C}{r} \quad (\text{A.1.20})$$

It therefore remains to check that $V_0^y(x) - \tilde{d}_0^y(x)/m$ satisfies the appropriate differential equation. To this end:

$$\begin{aligned}
& (1 - \phi)(x - C) + \tilde{d}_0^y(x) - p + \mathcal{L}_r^y(V_0^y - \tilde{d}_0^y/m) \\
&= -C + \tilde{d}_0^y - p - (1/m)\mathcal{L}_r^y\tilde{d}_0^y \\
&= -C + \tilde{d}_0^y(x) - p + (1/m)\left(c + mp - m\tilde{d}_0^y(x)\right) \\
&= 0,
\end{aligned} \tag{A.1.21}$$

where here I used the fact that $c + mp + \mathcal{L}_{r+m}^y\tilde{d}_0^y = 0$. This completes the proof. ■

Standard ODE techniques for the Cauchy-Euler equation give:

$$V_0^y(x) = U(x) + \frac{\phi C}{r} - \left[\xi U(x_{B,0}^y) + \frac{\phi C}{r} \right] \left(\frac{x}{x_{B,0}^y} \right)^{\gamma_2}, \tag{A.1.22}$$

where γ_2 is the negative root of the following polynomial equation:

$$g\gamma_2 + \frac{1}{2}y\gamma_2(\gamma_2 - 1) - r = 0. \tag{A.1.23}$$

The expression for $\tilde{d}_0^y(x)$ is provided in equation (1.4.11). Computing $V_0^y(x) - \tilde{d}_0^y(x)/m$ then gives equation (1.4.13). Differentiating this equation with respect to x and evaluating at $x_{B,0}^y$ to solve the smooth-pasting condition gives the endogenous default boundary (1.4.14).

A.1.3 Proof of Theorem 1.4.1

Let \bar{x}_B denote the exogenous default boundary. The problem for the principal order term $\tilde{d}_0^y(x)$ is the same as before, replacing $x_{B,0}^y$ with \bar{x}_B . Likewise, the solution to the problem is simply equation (1.4.11) with $x_{B,0}^y$ replaced by \bar{x}_B . I begin with a preliminary useful lemma.

Lemma A.1.5 *The following identity holds:*

$$\mathcal{L}_{r+m}^y \left\{ \frac{1}{g + \gamma_1 y - \frac{1}{2}y} \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \right\} = \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \tag{A.1.24}$$

Proof. By direct computation, I show that:

$$\begin{aligned}
& \mathcal{L}_{r+m}^y \left\{ \frac{1}{g + \gamma_1 y - \frac{1}{2}y} \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \right\} \\
&= \frac{1}{g + \gamma_1 y - \frac{1}{2}y} \left\{ \begin{aligned} & g \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \left[1 + \gamma_1 \ln \left(\frac{x}{\bar{x}_B} \right) \right] \\ & + \frac{1}{2}y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \left[2\gamma_1 - 1 + \gamma_1 (\gamma_1 - 1) \ln \left(\frac{x}{\bar{x}_B} \right) \right] \\ & - (r + m) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \ln \left(\frac{x}{\bar{x}_B} \right) \end{aligned} \right\} \\
&= \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \frac{g + \gamma_1 y - \frac{1}{2}y}{g + \gamma_1 y - \frac{1}{2}y} + \ln \left(\frac{x}{\bar{x}_B} \right) \frac{g\gamma_1 + \frac{1}{2}y\gamma_1 (\gamma_1 - 1) - (r + m)}{g + \gamma_1 y - \frac{1}{2}y} \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \\
&= \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1},
\end{aligned}$$

due the definition of γ_1 . ■

I next compute an explicit expression for the gamma and vega of the principal order term, that is the vega in the constant volatility model with exogenous default boundary.

Lemma A.1.6 *The gamma of the debt principal order term in the exogenous default model is given by:*

$$\frac{\partial^2 \tilde{d}_0^y}{\partial x^2} = \left[m(1 - \xi) U(\bar{x}_B) - \frac{c + mp}{r + m} \right] \gamma_1 (\gamma_1 - 1) \left\{ \frac{x}{\bar{x}_B} \right\}^{\gamma_1 - 2}. \quad (\text{A.1.25})$$

The vega of the principal order term is given by:

$$\frac{\partial \tilde{d}_0^y}{\partial y} = - \left[m(1 - \xi) U(\bar{x}_B) - \frac{c + mp}{r + m} \right] \frac{\gamma_1 (\gamma_1 - 1)}{2(g + \gamma_1 y - \frac{1}{2}y)} \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1}. \quad (\text{A.1.26})$$

Proof. The gamma of the principal order term can be calculated directly by differentiating equation (1.4.11) with respect to x . By the symmetry of partial derivatives, If I differentiate the ODE and boundary condition for \tilde{d}_0^y with respect to y , I get the following problem to solve for vega:

$$\mathcal{L}_{r+m}^y \frac{\partial \tilde{d}_0^y}{\partial y} + \frac{1}{2} x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2} = 0 \quad (\text{A.1.27a})$$

$$\frac{\partial \tilde{d}_0^y}{\partial y}(\bar{x}_B) = 0 \quad (\text{A.1.27b})$$

Since the gamma remains bounded as $x \rightarrow \bar{x}_B$, equation (A.1.26) clearly satisfies the boundary condition.

It remains to check that it satisfies the differential equation. This is a consequence of the lemma above.

Define:

$$A(\bar{x}_B) = m(1 - \xi) U(\bar{x}_B) - \frac{c + mp}{r + m}. \quad (\text{A.1.28})$$

Then I can show:

$$\begin{aligned}
& \mathcal{L}_{r+m}^y \left\{ -A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2(g + \gamma_1 y - \frac{1}{2}y)} \ln\left(\frac{x}{\bar{x}_B}\right) \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1} \right\} \\
= & -\frac{1}{2}A(\bar{x}_B) \gamma_1(\gamma_1 - 1) \mathcal{L}_{r+m}^y \frac{1}{g + \gamma_1 y - \frac{1}{2}y} \ln\left(\frac{x}{\bar{x}_B}\right) \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1} \tag{A.1.29}
\end{aligned}$$

$$\begin{aligned}
= & -\frac{1}{2}A(\bar{x}_B) \gamma_1(\gamma_1 - 1) \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1} \\
= & -\frac{1}{2}x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2}, \tag{A.1.30}
\end{aligned}$$

as desired. ■

Note that I could have simply calculated the vega of the principal order term directly through differentiation, recognizing that:

$$\begin{aligned}
\frac{\partial}{\partial y} \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1} &= \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1} \ln\left(\frac{x}{\bar{x}_B}\right) \frac{d\gamma_1}{dy} \\
&= -\frac{1}{2} \frac{\gamma_1(\gamma_1 - 1)}{g + \gamma_1 y - \frac{1}{2}y} \ln\left(\frac{x}{\bar{x}_B}\right) \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1}, \tag{A.1.31}
\end{aligned}$$

where $d\gamma_1/dy$ is computed through implicit differentiation of equation (1.4.12). However, the proof given above illustrates the usefulness of the first lemma, which I will now further exploit.

Lemma A.1.7 *The first-order debt correction term in the exogenous default model with $\rho_Y = 0$ is given explicitly by:*

$$\frac{\ln(x/\bar{x}_B)}{2(g + \gamma_1 y - \frac{1}{2}y)} A^\delta y \left[\frac{\partial \tilde{d}_0^y}{\partial y} + \frac{1}{2(g + \gamma_1 y - \frac{1}{2}y)^2} y x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2} \right] \tag{A.1.32}$$

Proof. I will break the calculation into parts. First, note that:

$$\begin{aligned}
& \mathcal{L}_{r+m}^y \frac{\ln(x/\bar{x}_B)}{2(g + \gamma_1 y - \frac{1}{2}y)} A^\delta y \frac{\partial \tilde{d}_0^y}{\partial y} \\
= & -\frac{1}{2}A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2(g + \gamma_1 y - \frac{1}{2}y)^2} A^\delta y \mathcal{L}_{r+m}^y \ln\left(\frac{x}{\bar{x}_B}\right)^2 \left(\frac{x}{\bar{x}_B}\right)^{\gamma_1} \tag{A.1.33}
\end{aligned}$$

Then:

$$\begin{aligned}
\mathcal{L}_{r+m}^y \ln \left(\frac{x}{\bar{x}_B} \right)^2 \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} &= g \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \left[2 \ln \left(\frac{x}{\bar{x}_B} \right) + \gamma_1 \ln \left(\frac{x}{\bar{x}_B} \right)^2 \right] \\
&\quad + \frac{1}{2} y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \left[2 + 2(2\gamma_1 - 1) \ln \left(\frac{x}{\bar{x}_B} \right) + \gamma_1 (\gamma_1 - 1) \ln \left(\frac{x}{\bar{x}_B} \right)^2 \right] \\
&\quad - (r+m) \ln \left(\frac{x}{\bar{x}_B} \right)^2 \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \\
&= y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} + 2 \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \left(g + \gamma_1 y - \frac{1}{2} y \right) \\
&\quad + \ln \left(\frac{x}{\bar{x}_B} \right)^2 \mathcal{L}_{r+m}^y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \\
&= y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} + 2 \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \left(g + \gamma_1 y - \frac{1}{2} y \right)
\end{aligned} \tag{A.1.34}$$

Therefore:

$$\begin{aligned}
&\mathcal{L}_{r+m}^y \frac{\ln(x/\bar{x}_B)}{2(g + \gamma_1 y - \frac{1}{2}y)} A^\delta y \frac{\partial \tilde{d}_0^y}{\partial y} \\
&= -\frac{1}{2} A^\delta y A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2} \frac{1}{(g + \gamma_1 y - \frac{1}{2}y)^2} y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \\
&\quad - A^\delta y A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2} \frac{1}{g + \gamma_1 y - \frac{1}{2}y} \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \\
&= -\frac{1}{2} A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2} A^\delta y \frac{1}{(g + \gamma_1 y - \frac{1}{2}y)^2} y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} + A^\delta y \frac{\partial \tilde{d}_0^y}{\partial y}
\end{aligned} \tag{A.1.35}$$

Finally, I show that:

$$\begin{aligned}
&\mathcal{L}_{r+m}^y \frac{\ln(x/\bar{x}_B)}{2(g + \gamma_1 y - \frac{1}{2}y)^3} A^\delta y \frac{1}{2} y x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2} \\
&= \frac{1}{2} A^\delta y A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2} \frac{1}{(g + \gamma_1 y - \frac{1}{2}y)^2} y \mathcal{L}_{r+m}^y \frac{1}{(g + \gamma_1 y - \frac{1}{2}y)} \ln \left(\frac{x}{\bar{x}_B} \right) \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1} \\
&= \frac{1}{2} A^\delta y A(\bar{x}_B) \frac{\gamma_1(\gamma_1 - 1)}{2} \frac{1}{(g + \gamma_1 y - \frac{1}{2}y)^2} y \left(\frac{x}{\bar{x}_B} \right)^{\gamma_1}.
\end{aligned} \tag{A.1.36}$$

This completes the proof since:

$$\mathcal{L}_{r+m}^y \frac{\ln(x/\bar{x}_B)}{2(g + \gamma_1 y - \frac{1}{2}y)} A^\delta y \left[\frac{\partial \tilde{d}_0^y}{\partial y} + \frac{1}{2(g + \gamma_1 y - \frac{1}{2}y)^2} y x^2 \frac{\partial^2 \tilde{d}_0^y}{\partial y^2} \right] = A^\delta y \frac{\partial \tilde{d}_0^y}{\partial y} \tag{A.1.37}$$

■

Note that an explicit expression can be derived for the case where $\rho_Y \neq 0$ as well. However, the computations and ultimate expression are quite messy and not very intuitive. I therefore omit them.

A.2 Contingent Claims of Growth Firms

In this appendix, I provide the problems to solve for the equity and debt values of growth firms and provide the details of how to solve for them by regular perturbation.

A.2.1 Value of Assets in Place

The value of assets in place for a growth firm is given by:

$$U_t^a(x) = \mathbb{E}_t^* \left[\int_t^\infty e^{-r(s-t)} (1-\phi) K_a X_s ds \mid X_t = x \right] = \frac{(1-\phi) K_a x}{r-g}. \quad (\text{A.2.1})$$

This is simply the Gordon growth formula for an unlevered firm with tax rate ϕ .

A.2.2 Equity Valuation of Growth Firms

The equity valuation of the growth firm is given by the following optimal stopping problem under the risk-neutral measure:

$$E_\delta^a(x, y) = \sup_{\tau', \tau'' \in \mathcal{T}} \mathbb{E}_t^* \left[\int_t^{\tau' \wedge \tau''} e^{-r(s-t)} \left\{ (1-\phi) (K_a X_s - C) + \tilde{d}_\delta^a(x, y) - p \right\} ds \right. \\ \left. + \mathbb{I}[\tau' < \tau''] (E_\delta(X_{\tau'}, Y_{\tau'}) - I) \right], \quad (\text{A.2.2})$$

where τ' is the stopping time which denotes investment and τ'' is the stopping time which denotes default. Upon defaulting, equityholders again receive nothing. At the investment threshold however, the equity value equals to the equity value a mature firm minus the cost of investment. Recall that by assumption, the cost of investment must be borne by equityholders, i.e., the firm cannot issue new debt to finance the purchase of additional capital.

By the same logic used to prove Theorem 1.3.1, we can characterize the equity valuation as a free boundary problem.

Theorem A.2.1 *The equity value of a growth firm $E_\delta^a(x, y)$ is the solution to:*

$$DIV_\delta^a(x, y) + \left(\mathcal{L}_r^y + \sqrt{\delta} \mathcal{M}_1^y + \delta \mathcal{M}_2^y \right) E_\delta^a = 0 \quad \text{for } x \in (x_{B,\delta}^a(y), x_{I,\delta}^a(y)) \quad (\text{A.2.3a})$$

$$E_\delta^a(x_{B,\delta}^a(y), y) = 0 \quad (\text{A.2.3b})$$

$$E_\delta^a(x_{I,\delta}^a(y), y) = E_\delta(x_{I,\delta}^a(y), y) - I \quad (\text{A.2.3c})$$

$$\frac{\partial E_\delta^a}{\partial x} \Big|_{x=x_{B,\delta}^a(y)} = 0 \quad (\text{A.2.3d})$$

$$\frac{\partial E_\delta^a}{\partial y} \Big|_{x=x_{B,\delta}^a(y)} = 0 \quad (\text{A.2.3e})$$

$$\frac{\partial E_\delta^a}{\partial x} \Big|_{x=x_{I,\delta}^a(y)} = \frac{\partial E_\delta}{\partial x} \Big|_{x=x_{I,\delta}^a(y)} \quad (\text{A.2.3f})$$

$$\frac{\partial E_\delta^a}{\partial y} \Big|_{x=x_{I,\delta}^a(y)} = \frac{\partial E_\delta}{\partial y} \Big|_{x=x_{I,\delta}^a(y)} \quad (\text{A.2.3g})$$

where

$$DIV_\delta^a(x, y) = (1 - \phi)(K_a x - C) + \tilde{d}_\delta^a(x, y) - p \quad (\text{A.2.4})$$

and $x_{B,\delta}^a(y)$ and $x_{I,\delta}^a(y)$ are free boundaries to be determined and

Proof. The proof follows the same logic as the proof of Theorem 1.3.1. ■

The free boundary problem is similar to the one specified for the equity value of mature firms. At the default boundary, the equity value must be equal to zero and the derivatives must be continuous. However, now there is no limiting condition as $x \rightarrow \infty$. Instead, there are additional value-matching and smooth-pasting conditions. At the investment threshold, the equity value must be equal to the equity of a mature firm minus the cost of investment. Once again, optimality in conjunction with the nature of a regular diffusion requires that the derivatives of the valuation be continuous across this barrier. Finally, the dividend to the equityholders now reflects the lower level of capital and the valuation of newly issued debt for young firms.

A.2.3 Debt Valuation of Growth Firms

Let τ^I denote the optimal stopping time for investment and τ^D the optimal stopping time for default:

$$\tau^I = \min \{t : (X_t, Y_t) = (x_{I,\delta}^a(Y_t), Y_t)\} \quad (\text{A.2.5})$$

$$\tau^B = \min \{t : (X_t, Y_t) = (x_{B,\delta}^a(Y_t), Y_t)\} \quad (\text{A.2.6})$$

Existing debt is simply rolled over when the firm invests in capital. Therefore, the value of a vintage of debt at time t issued at date 0 is given by:

$$d_{\delta}^a(t) = \mathbb{E}_t^* \left[\begin{aligned} & \int_t^{\tau^I \wedge \tau^D} \{e^{-r(s-t)} e^{-ms} (c + mp)\} ds \\ & + \mathbb{I}[\tau^D < \tau^I] e^{-r(\tau^D-t)} \left(e^{-m\tau^D} p/P \right) (1 - \xi) U_{\tau^D} \\ & + \mathbb{I}[\tau^I < \tau^D] e^{-r(\tau^I-t)} d_{\delta}(\tau^I) \end{aligned} \right] \quad (\text{A.2.7})$$

Upon default, debtholders receive a fraction of the value of the assets in place (minus bankruptcy costs) according to their vintage. At the investment threshold, the debt value equals the value of debt of identical vintage in a mature firm.

Noting that $p/P = m$ and $d_{\delta}(\tau^I) = e^{-m\tau^I} \tilde{d}_{\delta}(X_{\tau^I}, Y_{\tau^I})$, it follows that:

$$e^{mt} d_{\delta}^a(t) = \mathbb{E}_t^* \left[\begin{aligned} & \int_t^{\tau^I \wedge \tau^D} \{e^{-(r+m)(s-t)} (c + mp)\} ds \\ & + \mathbb{I}[\tau^D < \tau^I] e^{-(r+m)(\tau^D-t)} m (1 - \xi) U_{\tau^D}^a \\ & + \mathbb{I}[\tau^I < \tau^D] e^{-(r+m)(\tau^I-t)} \tilde{d}_{\delta}(X_{\tau^I}, Y_{\tau^I}) \end{aligned} \right], \quad (\text{A.2.8})$$

so that by Feynman-Kac the following theorem holds.

Theorem A.2.2 *The value of the date 0 debt vintage at time t for a young firm is given by $d^a(t) = e^{-mt} \tilde{d}^a(X_t, Y_t)$ where $\tilde{d}^a(X_t, Y_t)$ is the value of the newly issued debt and satisfies:*

$$c + mp + \left(\mathcal{L}_{r+m}^y + \sqrt{\delta} \mathcal{M}_1^y + \delta \mathcal{M}_2^y \right) \tilde{d}_{\delta}^a = 0 \quad \text{for } x \in (x_{B,\delta}^a(y), x_{I,\delta}^a(y)) \quad (\text{A.2.9a})$$

$$\tilde{d}_{\delta}^a(x_B^a(y), y) = m(1 - \xi) U^a(x_B^a(y)) \quad (\text{A.2.9b})$$

$$\tilde{d}_{\delta}^a(x_I^a(y), y) = \tilde{d}_{\delta}(x_I^a(y), y) \quad (\text{A.2.9c})$$

The total value of debt $D^a = \tilde{d}^a/m$.

I now expand the contingent claims and default boundaries in powers of $\sqrt{\delta}$:

$$E_{\delta}^a(x, y) = E_0^{a,y}(x) + \sqrt{\delta} E_1^{a,y}(x) + \delta E_2^{a,y}(x) + \dots \quad (\text{A.2.10a})$$

$$\tilde{d}_{\delta}^a(x, y) = \tilde{d}_0^{a,y}(x) + \sqrt{\delta} \tilde{d}_1^{a,y}(x) + \delta \tilde{d}_2^{a,y}(x) + \dots \quad (\text{A.2.10b})$$

$$x_B^a(y) = x_{B,0}^{a,y} + \sqrt{\delta} x_{B,1}^{a,y} + \delta x_{B,2}^{a,y} + \dots \quad (\text{A.2.10c})$$

These are substituted into the problems above. Boundary conditions are expanded with Taylor series.

Finally, contingent claims of mature firms are also written using asymptotic expansions.

A.2.4 Principal Order Terms

The principal order terms reflect the valuations and default boundary in the constant volatility case. The problem for debt is:

$$c + mp + \mathcal{L}_r^y \tilde{d}_0^{a,y} = 0 \quad \text{for } x \in (x_{B,0}^{a,y}, x_{I,0}^{a,y}) \quad (\text{A.2.11a})$$

$$\tilde{d}_0^{a,y} (x_{B,0}^{a,y}) = m (1 - \xi) U^a (x_{B,0}^{a,y}) \quad (\text{A.2.11b})$$

$$\tilde{d}_0^{a,y} (x_{I,0}^{a,y}) = \tilde{d}_0^y (x_{I,0}^{a,y}) \quad (\text{A.2.11c})$$

and the problem for equity is:

$$(1 - \phi) (xK_a - C) + \tilde{d}_0^{a,y} (x) - p + \mathcal{L}_r^y E_0^{a,y} = 0 \quad \text{or } x \in (x_{B,0}^{a,y}, x_{I,0}^{a,y}) \quad (\text{A.2.12a})$$

$$E_0^{a,y} (x_{B,0}^{a,y}) = 0 \quad (\text{A.2.12b})$$

$$E_0^{a,y} (x_{I,0}^{a,y}) = E_0^y (x_{I,0}^{a,y}) - I \quad (\text{A.2.12c})$$

$$\frac{\partial E_0^{a,y}}{\partial x} \Big|_{x=x_{B,0}^{a,y}} = 0 \quad (\text{A.2.12d})$$

$$\frac{\partial E_0^{a,y}}{\partial x} \Big|_{x=x_{I,0}^{a,y}} = \frac{\partial E_0^y}{\partial x} \Big|_{x=x_{I,0}^{a,y}} \quad (\text{A.2.12e})$$

I follow an indirect approach to solve for $E_0^{a,y}$ as in appendix A.1.2. I introduce the sum of debt and equity values $V_0^{a,y}$, which is the solution to:

$$(1 - \phi) K_a x + \phi C + \mathcal{L}_r^y V_0^{a,y} = 0 \quad \text{for } x \in (x_{B,0}^{a,y}, x_{I,0}^{a,y}) \quad (\text{A.2.13a})$$

$$V_0^{a,y} (x_{B,0}^{a,y}) = (1 - \xi) U^a (x_{B,0}^{a,y}) \quad (\text{A.2.13b})$$

$$V_0^{a,y} (x_{I,0}^{a,y}) = V_0^y (x_{I,0}^{a,y}) \quad (\text{A.2.13c})$$

Then $E_0^{a,y} = V_0^{a,y} - \tilde{d}_0^{a,y}/m$ and the default/investment boundaries are found by applying the two smooth-pasting conditions.

A.2.5 First-Order Correction Terms

Finally, the first-order corrections once again reflect comparative statics in the constant volatility case as well as boundary corrections. The first-order correction for debt is:

$$\mathcal{L}_{r+m}^y \sqrt{\delta} \tilde{d}_1^{a,y} = \left(A^\delta y \frac{\partial \tilde{d}_0^{a,y}}{\partial y} - B^\delta y x \frac{\partial^2 \tilde{d}_0^{a,y}}{\partial x \partial y} \right) \quad \text{for } x \in (x_{B,0}^{a,y}, x_{I,0}^{a,y}) \quad (\text{A.2.14a})$$

$$\sqrt{\delta} \tilde{d}_1^{a,y} (x_{B,0}^{a,y}) = \sqrt{\delta} x_{B,1}^{a,y} \left[\frac{m(1-\xi)(1-\phi)}{r-g} - \frac{\partial \tilde{d}_0^{a,y}}{\partial x} (x_{B,0}^{a,y}) \right] \quad (\text{A.2.14b})$$

$$\sqrt{\delta} \tilde{d}_1^{a,y} (x_{I,0}^{a,y}) = \sqrt{\delta} x_{I,1}^{a,y} \left[\frac{\partial \tilde{d}_0^y}{\partial x} (x_{I,0}^{a,y}) - \frac{\partial \tilde{d}_0^{a,y}}{\partial x} (x_{I,0}^{a,y}) \right] + \sqrt{\delta} \tilde{d}_1^y (x_{I,0}^{a,y}) \quad (\text{A.2.14c})$$

The first-order correction for equity is:

$$\mathcal{L}_r^y \sqrt{\delta} E_1^{a,y} = \left(A^\delta y \frac{\partial E_0^{a,y}}{\partial y} - B^\delta y x \frac{\partial^2 E_0^{a,y}}{\partial x \partial y} - \sqrt{\delta} \tilde{d}_1^{a,y} \right) \quad \text{for } x \in (x_{B,0}^{a,y}, x_{I,0}^{a,y}) \quad (\text{A.2.15a})$$

$$\sqrt{\delta} E_1^{a,y} (x_{B,0}^{a,y}) = 0 \quad (\text{A.2.15b})$$

$$\sqrt{\delta} E_1^{a,y} (x_{I,0}^{a,y}) = \sqrt{\delta} E_1^y (x_{I,0}^{a,y}) \quad (\text{A.2.15c})$$

Note that I used the smooth-pasting condition of the principal order term at the default and investment boundaries to derive the final two equations. Finally, the corrections in the default/investment barriers must satisfy:

$$\sqrt{\delta} x_{B,1}^{a,y} \frac{\partial^2 E_0^{a,y}}{\partial x^2} (x_{B,0}^{a,y}) = -\sqrt{\delta} \frac{\partial E_1^{a,y}}{\partial x} (x_{B,0}^{a,y}) \quad (\text{A.2.16})$$

$$\sqrt{\delta} x_{I,1}^{a,y} \frac{\partial^2 E_0^{a,y}}{\partial x^2} (x_{I,0}^{a,y}) = -\sqrt{\delta} \frac{\partial E_1^{a,y}}{\partial x} (x_{I,0}^{a,y}) \quad (\text{A.2.17})$$

All of this together solves a system of equations which can be solved numerically in MATLAB.

A.3 Volatility Time Scales

This technical appendix elaborates on the discussion of volatility time scales in section 1.4.1 and specifically on the role of δ in controlling the rate of mean-reversion. The discussion follows closely the textbook treatment provided in Fouque, Papanicolaou, Sircar, and Solna (2011). I define the infinitesimal generator of a time-homogenous, ergodic Markov process Y_t to be:

$$\mathcal{L}h(y) = \lim_{t \rightarrow 0} \frac{\mathbf{P}_t h(y) - h(y)}{t}, \quad (\text{A.3.1})$$

where:

$$\mathbf{P}_t h(y) = \mathbb{E}[h(Y_t) \mid Y_0 = y]. \quad (\text{A.3.2})$$

For example, the infinitesimal generator of the Cox-Ingersoll-Ross process of equation (1.3.2) is given by:

$$\mathcal{M}_2^y = (\theta_Y - y) \frac{\partial}{\partial y} + \frac{1}{2} \nu_Y^2 y \frac{\partial^2}{\partial y^2}. \quad (\text{A.3.3})$$

In general, the infinitesimal generator of a regular diffusion can be found by considering Ito's formula and the backwards Kolmogorov equation. To find the invariant distribution of the process Y_t , which exists by ergodicity, I look for a distribution Λ for Y_0 which satisfies for any bounded h :

$$\frac{d}{dt} \int \mathbb{E}[h(Y_t) \mid Y_0 = y] d\Lambda(y) = 0, \quad (\text{A.3.4})$$

where the integral is taken of the state space on which the Markov process is defined. By the backward Kolmogorov equation for a time-homogenous Markov process:

$$\frac{d}{dt} \mathbf{P}_t h(y) = \mathcal{L} \mathbf{P}_t h(y), \quad (\text{A.3.5})$$

it follows that:

$$\begin{aligned} \frac{d}{dt} \int \mathbb{E}[h(Y_t) \mid Y_0 = y] d\Lambda(y) &= \int \frac{d}{dt} \mathbf{P}_t h(y) d\Lambda(y) \\ &= \int \mathcal{L} \mathbf{P}_t h(y) d\Lambda(y) \\ &= \int \mathbf{P}_t h(y) \mathcal{L}^* d\Lambda(y), \end{aligned} \quad (\text{A.3.6})$$

where \mathcal{L}^* is the adjoint operator of \mathcal{L} defined uniquely by:

$$\int \alpha(y) \mathcal{L} \beta(y) dy = \int \beta(y) \mathcal{L}^* \alpha(y) dy \quad (\text{A.3.7})$$

for test functions α, β . The invariant distribution, therefore, is the solution to the equation:

$$\mathcal{L}^* \Lambda = 0, \quad (\text{A.3.8})$$

since the relation above must hold for all h . I denote integration with respect to the invariant distribution by $\langle h \rangle$ and suppose that the invariant distribution has a mean.

Now define the process Y_t^δ according the infinitesimal generator $\delta\mathcal{L}$. For a CIR process, this procedure then gives a new process which is given explicitly in equation (1.4.1). The adjoint of the operator $\delta\mathcal{L}$ is clearly given by $\delta\mathcal{L}^*$. Therefore, it is immediately clear that the invariant distribution of the process Y_t^δ is independent of the choice of δ , as described in section 1.4.1. This indicates that, in the long-run, the choice of δ does not affect the degree of variability in the process Y_t^δ .

Now I suppose that the process Y_t is reversible, or that the operator \mathcal{L} has a discrete spectrum and that zero is an isolated eigenvalue. This allows the formation of an orthonormal basis of $L^2(\Lambda)$ and to consider the eigenfunction expansion of a function $h(y)$ by:

$$h(y) = \sum_{n=0}^{\infty} d_n \psi_n(y) \quad (\text{A.3.9})$$

where each ψ_n satisfies:

$$\mathcal{L}\psi_n = \lambda_n \psi_n \quad (\text{A.3.10})$$

and where $0 = \lambda_0 > \lambda_1 > \dots$. Each constant satisfies $d_n = \langle h \psi_n \rangle$ and in particular $d_0 = \langle h \rangle$, since the eigenfunction associated with the zero eigenvalue is simply $\psi_0 = 1$. Next consider the backwards Kolmogorov equation:

$$\frac{d}{dt} \mathbf{P}_t h(y) = \mathcal{L} \mathbf{P}_t h(y) \quad (\text{A.3.11})$$

and look for a solution of the form:

$$\mathbf{P}_t h(y) = \sum_{n=0}^{\infty} z_n(t) \psi_n(y). \quad (\text{A.3.12})$$

Substituting this expression into the backwards Kolmogorov equation and using equation (A.3.10) gives an ODE for each $z_n(t)$:

$$z'_n(t) = \lambda_n z_n(t), \quad (\text{A.3.13})$$

with initial condition $z_n(0) = d_n$. Solving, this implies that:

$$\mathbf{P}_t h(y) = \sum_{n=0}^{\infty} c_n e^{\lambda_n t} \psi_n(y). \quad (\text{A.3.14})$$

From this, it is possible to show that:

$$|\mathbf{P}_t g(y) - \langle h \rangle| \leq C e^{\lambda_1 t}, \quad (\text{A.3.15})$$

for some constant C . In words, the spectral gap, defined as the magnitude of the first negative eigenvalue,

controls the rate of mean reversion of the process to the invariant distribution.

Finally, it is trivial to see that the eigenfunctions/eigenvalues of the infinitesimal generator $\delta\mathcal{L}$ are given by:

$$\mathcal{L}\delta\psi_n = \lambda_n\delta\psi_n. \tag{A.3.16}$$

That is, the spectrum of $\delta\mathcal{L}$ is simply a scaling of the spectrum of \mathcal{L} according to the parameter δ . But this implies that the spectral gap of the process Y_t^δ is proportional to δ and, therefore, that δ controls the rate of mean-reversion of the process.

Appendix B

Supplement to Chapter 2

B.1 Derivations and Proofs

B.1.1 Steady State

Denoting steady state values with a star superscript, the no default assumption means that $l_1^* = 0$, $v_d^* = 0$, $v_r^* = 0$, $v_n^* = v_b^*$ and $l_0^* = 1 - v_n^*$. This implies $\mu^* = 1$, $q_s(\mu^*) = q_b(\mu^*) = \chi$, $r_n^* = 1$, and $r_d^* = 0$. Because inflows into being a seller and outflows from being a seller are equal in steady state,

$$\begin{aligned} (1 - v_n^*)\gamma &= v_n^*\chi(1 - F(h_n^*)) \\ v_n^* &= \frac{\gamma}{\gamma + \chi(1 - F(h_n^*))} \quad l_0 = \frac{\chi(1 - F(h_n^*))}{\gamma + \chi(1 - F(h_n^*))}. \end{aligned} \quad (\text{B.1.1})$$

Replacing the conditional expectations of surpluses with differences of cutoffs as in (2.3.12) and setting $r_n^* = 1$, $r_d^* = 0$, and $q_s^* = q_b^* = \chi$ yields simplified steady state value functions:

$$V_h^* = \frac{h + \beta\{\gamma(V_n^* + B^*)\}}{1 - \beta(1 - \gamma)} \quad (\text{B.1.2})$$

$$B^* = \frac{u_b}{1 - \beta} + \frac{(1 - \theta)\chi(1 - F(h_n^*))E[h - h_n^*|h \geq h_n^*]}{(1 - \beta)(1 - \beta(1 - \gamma))} \quad (\text{B.1.3})$$

$$V_m^* = \frac{m}{1 - \beta} + \frac{\theta\chi(1 - F(h_m^*))E[h - h_m^*|h \geq h_m^*]}{(1 - \beta)(1 - \beta(1 - \gamma))}, m \in \{n, d\} \quad (\text{B.1.4})$$

$$R^* = \frac{u_r + \beta\sigma B^*}{1 - \beta(1 - \sigma)}. \quad (\text{B.1.5})$$

With everything in terms of the cutoffs, a two-equation system that pins down h_n^* and h_d^* . Subtracting the cutoff condition (2.3.7) at the distressed and non-distressed cutoffs gives:

$$V_{h_n}^* - V_{h_d}^* = (m_n - m_d) + \beta [V_n^* - V_d^*]. \quad (\text{B.1.6})$$

Plugging in the steady state values and manipulating yields an equation that implicitly defines the difference of the cutoffs:

$$\begin{aligned} h_n^* - h_d^* &= (m_n - m_d) \frac{1 - \beta(1 - \gamma)}{(1 - \beta)} \\ &+ \theta \frac{\beta}{1 - \beta} \chi \{ (1 - F(h_n^*)) E[h - h_n^* | h \geq h_n^*] - (1 - F(h_d^*)) E[h - h_d^* | h \geq h_d^*] \}. \end{aligned} \quad (\text{B.1.7})$$

The second equation comes from evaluating the cutoff condition (2.3.7) at h_n^* :

$$V_{h_n}^* = m_n + u_b + \beta [B^* + V_n^*]. \quad (\text{B.1.8})$$

Equations (B.1.7) and (B.1.8) define a system that can be solved for h_d^* and h_n^* . All of the other steady-state variables are written in terms of these cutoffs.

Theorem B.1.1 *If $a < m_n + u_b + \frac{\beta(1-\gamma)\chi E[h-a]}{1-\beta(1-\gamma)}$, there exists a unique steady state of the model.*

Proof. Because there are no REO sellers in steady state,

$$V_n^* + B^* = \frac{m_n + u_b}{1 - \beta} + \frac{\chi(1 - F(h_n^*)) E[h - h_n^* | h \geq h_n^*]}{(1 - \beta)(1 - \beta(1 - \gamma))} \quad (\text{B.1.9})$$

The cutoff condition for h_n^* is:

$$\begin{aligned} \frac{h_n^* + \beta[V_n^* + B^*]}{1 - \beta(1 - \gamma)} &= m_n + u_b + \beta[V_n^* + B^*] \\ \frac{h_n^*}{1 - \beta(1 - \gamma)} &= m_n + u_b + \beta \left(\frac{1 - \beta(1 - \gamma) - \gamma}{1 - \beta(1 - \gamma)} \right) [V_n^* + B^*] \\ h_n^* &= (1 - \beta(1 - \gamma))(m_n + u_b) + \beta(1 - \beta)(1 - \gamma)[V_n^* + B^*] \end{aligned} \quad (\text{B.1.10})$$

Plugging in for $V_n^* + B^*$ and re-arranging yields:

$$\begin{aligned} h_n^* &= (1 - \beta(1 - \gamma))(m_n + u_b) + \beta(1 - \beta)(1 - \gamma) \left[\frac{m_n + u_b}{1 - \beta} + \frac{\chi(1 - F(h_n^*)) E[h - h_n^* | h \geq h_n^*]}{(1 - \beta)(1 - \beta(1 - \gamma))} \right] \\ h_n^* &= m_n + u_b + \frac{\beta(1 - \gamma)\chi(1 - F(h_n^*)) E[h - h_n^* | h \geq h_n^*]}{1 - \beta(1 - \gamma)} \end{aligned} \quad (\text{B.1.11})$$

We want to find a unique solution to this equation on $h_n^* \in [a, \infty)$. As $h_n^* \rightarrow \infty$, the RHS of equation (B.1.11) approaches $m_n - u_b$. As $h_n^* \rightarrow a$, the RHS of equation (B.1.11) approaches $m_n + u_b + \frac{\beta(1-\gamma)\chi E[h-a]}{1-\beta(1-\gamma)}$. Thus as long as $a < m_n + u_b + \frac{\beta(1-\gamma)\chi E[h-a]}{1-\beta(1-\gamma)}$, since both the RHS and LHS of equation (B.1.11) are continuous in h_n^* , by the intermediate value theorem we know there exists a solution on $[a, \infty)$ to equation (B.1.11). Furthermore, we know that the LHS is strictly increasing in h_n^* , while the RHS is strictly decreasing in h_n^* since:

$$\frac{d}{dx} (1 - F(x)) E[h - x | h \geq x] = -(1 - F(x)) < 0. \quad (\text{B.1.12})$$

This implies that the solution to equation (B.1.11) is unique.

Finally note that

$$h_d^* - \theta \frac{\beta}{1-\beta} \chi (1 - F(h_d^*)) E[h - h_d^* | h \geq h_d^*] \quad (\text{B.1.13})$$

is a monotonically increasing function of h_d^* and thus, given the solution h_n^* to equation (B.1.11), there exists a unique solution to equation (B.1.7). ■

Note that given our assumptions it is generally the case that $a < m_n + u_b + \frac{\beta(1-\gamma)\chi E[h-a]}{1-\beta(1-\gamma)}$ and thus that there is a unique equilibrium. This is because β is close to 1 and γ is close to 0, so the denominator of the fraction is very small. Uniqueness would only be a concern with a very low discount factor or high moving probability.

Due to higher holding costs and balance sheet concerns, an REO seller should be more willing to sell the property conditional on being matched with a buyer than a normal seller. We show this is always the case in steady state:

Lemma B.1.2 *For a given h , the probability of sale for a distressed seller is higher than the probability of sale for a non-distressed seller.*

Proof. Note again that:

$$\frac{d}{dx} (1 - F(x)) E[h - x | h \geq x] = -(1 - F(x)) < 0. \quad (\text{B.1.14})$$

Suppose that $h_n^* \leq h_d^*$. Then:

$$\begin{aligned} (h_n^* - h_d^*) & - \theta \frac{\beta}{1-\beta} \chi \{ (1 - F(h_n^*)) E[h - h_n^* | h \geq h_n^*] - (1 - F(h_d^*)) E[h - h_d^* | h \geq h_d^*] \} \\ & < 0 \\ & < (m_n - m_d) \frac{1 - \beta(1 - \gamma)}{(1 - \beta)}. \end{aligned} \quad (\text{B.1.15})$$

which contradicts equation (B.1.7). It must therefore be that $h_n^* > h_d^*$, which indicates that distressed sellers are more likely to sell than non-distressed sellers. ■

We use the Nash bargaining condition to back out steady state prices. We have for $m \in \{n, d\}$ and a given h :

$$\frac{\theta}{1-\theta} = \frac{S_m^S}{S_m^B} = \frac{p_m^*(h) - m - \beta V_m^*}{V_h^* - p_m^*(h) - \beta B^*} \quad (\text{B.1.16})$$

And so

$$\begin{aligned} \theta [V_h^* - p_m^*(h) - \beta B^*] &= (1-\theta) [p_m^*(h) - m - \beta V_m^*] \\ p_m^*(h) &= \theta V_h^* - \beta \theta B^* + (1-\theta) m + (1-\theta) \beta V_m^* \\ &= \theta V_h^* + m + \beta V_m^* - \theta [m + \beta V_m^* + \beta B^*] \\ &= \theta [V_h^* - V_{h_m^*}^*] + m + \beta V_m^* \\ &= \frac{\theta (h - h_m^*)}{1 - \beta (1 - \gamma)} + m + \frac{\beta m}{1 - \beta} + \frac{\beta \theta \chi (1 - F(h_m^*)) E[h - h_m^* | h \geq h_m^*]}{(1 - \beta) (1 - \beta (1 - \gamma))} \\ &= \frac{\theta (h - h_m^*)}{1 - \beta (1 - \gamma)} + \frac{m}{1 - \beta} + \frac{\beta \theta \chi (1 - F(h_m^*)) E[h - h_m^* | h \geq h_m^*]}{(1 - \beta) (1 - \beta (1 - \gamma))} \quad (\text{B.1.17}) \end{aligned}$$

It is also always the case that distressed properties sell for less than non-distressed properties:

Theorem B.1.3 *Distressed sales trade at a constant discount, in the sense that $p_n^*(h) - p_d^*(h) = \Delta$ for all $h \geq h_n^* > h_d^*$ for some constant $\Delta > 0$.*

Proof. Taking the difference of prices we get:

$$\begin{aligned} p_n^*(h) - p_d^*(h) &= \frac{\theta (h_d^* - h_n^*)}{1 - \beta (1 - \gamma)} + \frac{m_n - m_d}{1 - \beta} \\ &\quad + \frac{\beta \theta \chi \{(1 - F(h_n^*)) E[h - h_n^* | h \geq h_n^*] - (1 - F(h_d^*)) E[h - h_d^* | h \geq h_d^*]\}}{(1 - \beta) (1 - \beta (1 - \gamma))} \end{aligned}$$

Using equation (B.1.7) this simplifies to:

$$\begin{aligned} p_n(h) - p_d(h) &= \frac{\theta (h_d^* - h_n^*)}{1 - \beta (1 - \gamma)} + \frac{m_n - m_d}{1 - \beta} - \frac{m_n - m_d}{1 - \beta} + \frac{h_n^* - h_d^*}{1 - \beta (1 - \gamma)} \\ &= \frac{1 - \theta}{1 - \beta (1 - \gamma)} (h_n^* - h_d^*) \equiv \Delta > 0 \quad (\text{B.1.18}) \end{aligned}$$

■

B.1.2 Dynamics

This appendix describes how we solve the dynamic model with exogenous defaults presented in section 2.3.

First, the laws of motion simply add inflows and subtract outflows according to figure 2.3.2:

$$l_0(t+1) = (1-\gamma)l_0(t) + v_b(t)q_b(\mu(t)) \sum_m r_m(t) e^{-\lambda(h_m(t)-a)} \quad (\text{B.1.19})$$

$$l_1(t+1) = (1-\gamma)l_1(t) \quad (\text{B.1.20})$$

$$v_n(t+1) = \gamma l_0(t) + \gamma(1-\alpha)l_1(t) + v_n \left[1 - q_s(\mu(t)) e^{-\lambda(h_n(t)-a)} \right] \quad (\text{B.1.21})$$

$$v_d(t+1) = \gamma \alpha l_1(t) + v_d \left[1 - q_s(\mu(t)) e^{-\lambda(h_d(t)-a)} \right] \quad (\text{B.1.22})$$

$$v_b(t+1) = \gamma l_0(t) + \gamma(1-\alpha)l_1(t) + v_r(t)\sigma + v_b(t) \left[1 - q_b(\mu(t)) \sum_m r_m(t) e^{-\lambda(h_m(t)-a)} \right] \quad (\text{B.1.23})$$

$$v_r(t+1) = \gamma \alpha l_1(t) + (1-\sigma)v_r(t) \quad (\text{B.1.24})$$

To generate the dynamic path of V_h , we expand the sum in equation (2.3.2) and collecting terms:

$$V_h(t) = \frac{h}{1-\beta(1-\gamma)} + \beta \sum_{j=1}^{\infty} \left[(\beta(1-\gamma))^{j-1} \{ \gamma [V_n(t+j) + B(t+j)] \} \right]. \quad (\text{B.1.25})$$

The sum in the second term can be written recursively as:

$$\Gamma(t) = \gamma [V_n(t) + B(t)] + \beta(1-\gamma)\Gamma(t+1) \quad (\text{B.1.26})$$

so that

$$V_h(t) = \frac{h}{1-\beta(1-\gamma)} + \beta \Gamma(t+1). \quad (\text{B.1.27})$$

The entire dynamic system can thus be written recursively. The cutoff rule then simplifies to:

$$\frac{h_m(t)}{1-\beta(1-\gamma)} + \beta \Gamma(t+1) = u_b + \beta B(t+1) + m + \beta V_m(t+1). \quad (\text{B.1.28})$$

The full general equilibrium model is made up of 13 endogenous variables and 13 equations. The endogenous variables are the cutoffs h_n and h_d , the masses v_n , v_d , v_b , v_r , l_0 , and l_1 , and value functions Γ , V_{m_d} , V_{m_r} , B , and R . From these values, all of the other endogenous parameters of the model can be determined. Substituting out the conditional expectation of the surplus using (2.3.12), using the exponential

distribution, and using the definitions of q_b and q_s from the matching function gives the dynamic system:

$$\begin{aligned}
V_n(t) &= \beta V_n(t+1) + m_n + \chi\mu(t) \frac{\theta}{\lambda[1-\beta(1-\gamma)]} e^{-\lambda(h_n(t)-a)} \\
V_d(t) &= \beta V_d(t+1) + m_d + \chi\mu(t) \frac{\theta}{\lambda[1-\beta(1-\gamma)]} e^{-\lambda(h_d(t)-a)} \\
B(t) &= \beta B(t+1) + u_b + \chi\mu(t) \frac{1-\theta}{\lambda[1-\beta(1-\gamma)]} \sum_m r_m(t) e^{-\lambda(h_m(t)-a)} \\
R(t) &= u_r + \beta \{ \sigma B(t+1) + (1-\sigma)R(t+1) \} \\
\Gamma(t) &= \gamma [V_n(t) + B(t)] + \beta(1-\gamma)\Gamma(t+1) \\
l_0(t+1) &= (1-\gamma)l_0(t) + v_b(t)\chi\mu(t) \sum_m r_m(t) e^{-\lambda(h_m(t)-a)} \\
l_1(t+1) &= (1-\gamma)l_1(t) \\
v_n(t+1) &= \gamma l_0(t) + \gamma(1-\alpha)l_1(t) + v_n \left[1 - \chi\mu(t) e^{-\lambda(h_n(t)-a)} \right] \\
v_d(t+1) &= \gamma \alpha l_1(t) + v_d \left[1 - \chi\mu(t) e^{-\lambda(h_d(t)-a)} \right] \\
v_b(t+1) &= \gamma l_0(t) + \gamma(1-\alpha)l_1(t) + v_r(t)\sigma + v_b(t) \left[1 - \chi\mu(t) \sum_m r_m(t) e^{-\lambda(h_m(t)-a)} \right] \\
v_r(t+1) &= \gamma \alpha l_1(t) + (1-\sigma)v_r(t) \\
\frac{h_n(t)}{1-\beta(1-\gamma)} + \beta\Gamma(t+1) &= u_b + \beta B(t+1) + m_n + \beta V_n(t+1) \\
\frac{h_d(t)}{1-\beta(1-\gamma)} + \beta\Gamma(t+1) &= u_b + \beta B(t+1) + m_d + \beta V_d(t+1)
\end{aligned}$$

where $\mu(t) = \frac{v_b(t)}{v_n(t)+v_d(t)}$ and $r_m(t) = \frac{v_m(t)}{v_n(t)+v_d(t)}$. We solve this system using Newton's Method as implemented in DYNARE, which guesses that the model returns to steady state at time T , solves a system of $13T$ equations, and checks that the model is in fact within ε of the steady state at time T . Solving the model with endogenous defaults of section 2.5 is performed similarly although the laws of motion are modified as described in appendix B.1.4.

We then back out prices as in the steady state. Prices are defined by (2.3.13). The mean price for a type m seller is then:

$$\bar{p}_m = E_h [p_{m,h}(t) | h_m \geq h_m] = \frac{\theta}{\lambda} \frac{1}{1-\beta(1-\gamma)} + m + \beta V_m(t+1) \quad (\text{B.1.29})$$

and the overall mean price in a price index is:

$$\begin{aligned}
\bar{p} &= FVol_N \left[\frac{\theta}{\lambda} \frac{1}{1-\beta(1-\gamma)} + m_n + \beta V_n(t+1) \right] \\
&\quad + FVol_D \left[\frac{\theta}{\lambda} \frac{1}{1-\beta(1-\gamma)} + m_d + \beta V_d(t+1) \right]
\end{aligned} \quad (\text{B.1.30})$$

where $FVol_m$ is the fraction of total volume accounted for by type m sellers.

B.1.3 Calibration

As described in the main text, we use five aggregate moments from the housing market prior to the crash to set a , λ , m_n , and m_d . These moments are the average price of a normal home in steady state \bar{p}_n^* , the variance of the residual price distribution $\sigma_{p_n^*}^2$, the discount for a distressed sale in terms of mean prices $\frac{\bar{p}_n - \bar{p}_d}{\bar{p}_n}$, and the time on the market for a normal sale T_n^* and a distressed sale T_d^*

Using the expressions for the price and the probability of sale in the main text along with properties of the exponential distribution, these moments are:

$$\bar{p}_n^* = \frac{\theta}{\lambda [1 - \beta (1 - \gamma)]} + \frac{m_n}{1 - \beta} + \frac{\beta \theta \chi e^{-\lambda (h_n^* - a)}}{\lambda (1 - \beta) (1 - \beta (1 - \gamma))} \quad (\text{B.1.31})$$

$$\sigma_{p_n^*}^2 = \frac{\theta^2}{\lambda^2 [1 - \beta (1 - \gamma)]^2} \quad (\text{B.1.32})$$

$$\frac{\bar{p}_n - \bar{p}_d}{\bar{p}_n} = \frac{m_n - m_d}{\bar{p}_n (1 - \beta)} + \frac{\beta \theta \chi \{e^{-\lambda (h_n^* - a)} - e^{-\lambda (h_d^* - a)}\}}{\lambda \bar{p}_n (1 - \beta) (1 - \beta (1 - \gamma))} = \frac{h_n^* - h_d^*}{\bar{p}_n [1 - \beta (1 - \gamma)]} \quad (\text{B.1.33})$$

$$T_n^* = \frac{1 - \chi \exp(-\lambda (h_n^* - a))}{\chi \exp(-\lambda (h_n^* - a))} = \frac{1}{\chi} \exp(\lambda (h_n^* - a)) - 1 \quad (\text{B.1.34})$$

$$T_d^* = \frac{1 - \chi \exp(-\lambda (h_d^* - a))}{\chi \exp(-\lambda (h_d^* - a))} = \frac{1}{\chi} \exp(\lambda (h_d^* - a)) - 1 \quad (\text{B.1.35})$$

Plugging the second and fourth equations into the first gives:

$$\bar{p}_n^* = \sigma_{p_n^*} + \frac{m_n}{1 - \beta} + \frac{\beta}{1 - \beta} \frac{\sigma_{p_n^*}}{T_n^* + 1} \quad (\text{B.1.36})$$

which implicitly defines m_n as a function of known parameters and observable moments.

We then define a six equation system with six variables – a , θ , λ , m_d , h_n^* , and h_d^* – that we use to calibrate the remainder of the model. Taking the square root of the second equation and rearranging gives λ as a function of θ and observable moments:

$$\lambda = \frac{\theta}{\sigma_{p_n^*} [1 - \beta (1 - \gamma)]}. \quad (\text{B.1.37})$$

An expression for a is obtained by inverting the fourth equation $h_n^* = a + \frac{1}{\lambda} \ln(\chi (T_n^* + 1))$ and then plugging

into the cutoff condition for n :

$$V_{h_n^*} = m_n + u_b + \beta [B^* + V_{m_d}^*] \quad (\text{B.1.38})$$

$$h_n^* = (1 - \beta(1 - \gamma)) m_n + \beta(1 - \gamma) \left[m_n + \frac{\chi e^{-\lambda(h_n^* - a)}}{\lambda(1 - \beta(1 - \gamma))} \right] \quad (\text{B.1.39})$$

Plugging in for h_n^* and solving gives:

$$a = (1 - \beta(1 - \gamma)) m_n + \beta(1 - \gamma) \left[m_n + \frac{1}{(T_n^* + 1) \lambda(1 - \beta(1 - \gamma))} \right] - \frac{1}{\lambda} \ln(\chi(T_n^* + 1)) \quad (\text{B.1.40})$$

The equations for λ and a , the moments for T_n^* , T_d^* , and the discount,

$$\frac{\bar{p}_n - \bar{p}_d}{\bar{p}_n} = \frac{h_n^* - h_d^*}{\bar{p}_n [1 - \beta(1 - \gamma)]} \quad T_n^* = \frac{1}{\chi} \exp(\lambda(h_n^* - a)) - 1 \quad T_d^* = \frac{1}{\chi} \exp(\lambda(h_d^* - a)) - 1, \quad (\text{B.1.41})$$

along with (B.1.7),

$$h_n^* - h_d^* = (m_n - m_d) \frac{1 - \beta(1 - \gamma)}{(1 - \beta)} + \frac{\theta}{\lambda} \frac{\beta}{1 - \beta} \chi \left\{ e^{-\lambda(h_n^* - a)} - e^{-\lambda(h_d^* - a)} \right\}, \quad (\text{B.1.42})$$

form the six equation system, which we solve numerically.

Although all of the variables are jointly determined, we have found that θ and the gap between m_d and m_n are principally determined by the gap in time on the market and the REO discount while a and λ are principally determined by the moments of the price distribution.

B.1.4 Extended Model

For the extended model, the housing market is unchanged and so the value functions are unchanged. Only the laws of motion differ, as described by Figure 2.5.1. As described in the text, we have two different exogenous shocks. First, we assume that a fraction $\delta(t)$ of individuals who sell due to taste shocks become

renters instead of buyers and shock $\delta(t)$. This leads to the following laws of motion:

$$\begin{aligned}
l_0(t+1) &= (1-\gamma)l_0(t) + v_b(t)q_b(\mu(t))\sum r_m(t)(1-F(h_m(t))) + w(t)\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))} \\
l_1(t+1) &= (1-\gamma)l_1(t) \\
w(t+1) &= (1-\gamma_I)w(t) + (\gamma-\gamma_I)l_1(t)(1-G(V_n(t))) - w(t)\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))} \\
f(t+1) &= \gamma_I l_1(t)(1-G(V_n(t))) + \gamma_I w(t) \\
&\quad + f(t)\left(1 - \frac{1}{\phi f(t)+1}\right) - f(t)\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))} \\
v_n(t+1) &= \gamma l_0(t) + \gamma l_1(t)G(V_n(t)) + (w(t) + f(t))\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))} \\
&\quad + v_n[1 - q_s(\mu(t))(1-F(h_n(t)))] \\
v_d(t+1) &= \frac{f(t)}{\phi f(t)+1} + v_d[1 - q_s(\mu(t))(1-F(h_d(t)))] \\
v_b(t+1) &= (1-\delta)[\gamma l_0(t) + \gamma l_1(t)G(V_n(t))] + \\
&\quad + v_r(t)\sigma + v_b(t)\left[1 - q_b(\mu(t))\sum r_m(t)(1-F(h_m(t)))\right] \\
v_r(t+1) &= \delta\left[\gamma_I l_0(t) + \gamma_I l_1(t)G(V_n(t)) + w(t)\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))}\right] \\
&\quad + f(t)\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))} + \frac{f(t)}{\phi f(t)+1} + (1-\sigma)v_r(t)
\end{aligned}$$

Second, we assume that a falls permanently and that γ_I declines gradually after 10 years. This is the same as setting $\delta = 0$ above, shocking a , and using the following auto-regressive process for a :

$$\gamma_I = \tau_t \bar{\gamma}_I \text{ where } \tau_t = \alpha \tau_{t-1} \text{ and } \tau_1 = 1$$

We assume $\alpha = 1$ for five years when it falls to $\alpha = .95$.

These laws of motion simply add inflows and subtract outflows. A fraction $(1 - G(V_n(t)))$ of individuals who receive taste shocks default and the same fraction of individuals with taste shocks become locked in. A fraction δ of individuals who would become buyers and sellers become a buyer and a renter instead. A mass $\frac{f(t)}{\phi f(t)+1}$ experiences a foreclosure completion. The final added complexity is accounting for the mass of individuals who were locked in in period $t-1$ but are no longer locked in in period t or who were in foreclosure in period $t-1$ but are no longer in foreclosure in period t due to rising prices. Because only individuals with a loan balance above $V_n(t-1)$ are locked in at time $t-1$, this mass is a fraction $\frac{G(V_n(t))-G(V_n(t-1))}{1-G(V_n(t-1))}$ of the mass $w(t)$ and $f(t)$, respectively.

These laws of motion replace the laws of motion in appendix B.1.2 above. The rest of the equations are the same, yielding a 15 equation and 15 unknown dynamic system.

B.2 Details Omitted From Main Text

B.2.1 Details of Isolating Each Effect

As described in the main text we perform three experiments to isolate the role of each driving force in our model. We provide the details of these experiments here:

1. To shut down the market tightness effect, we assume that a homeowner who defaults is not forced to become a renter for a certain random amount of time, but can instead immediately re-enter the housing market as a buyer. The law of motion for the stock of buyers in the market then becomes

$$v_b(t+1) = \gamma(l_0(t) + l_1(t)) + v_b(t) \left[1 - q_b(\mu) \sum r_m(t) e^{-\lambda(h_m(t)-a)} \right], \quad (\text{B.2.1})$$

while all other equations remain unchanged. Note that the stock of renters will always be zero in this experiment and $\mu(t) = 1$ for all t .

2. To shut down the choosey buyer effect, we assume that the buyer believes every seller he meets will be a retail seller. That is, even though there may well be distressed sellers in the market, the buyer fails to take their presence into account when determining his optimal market behavior. Along these lines, for this experiment we modify the Bellman equation of the buyer's value function to read:

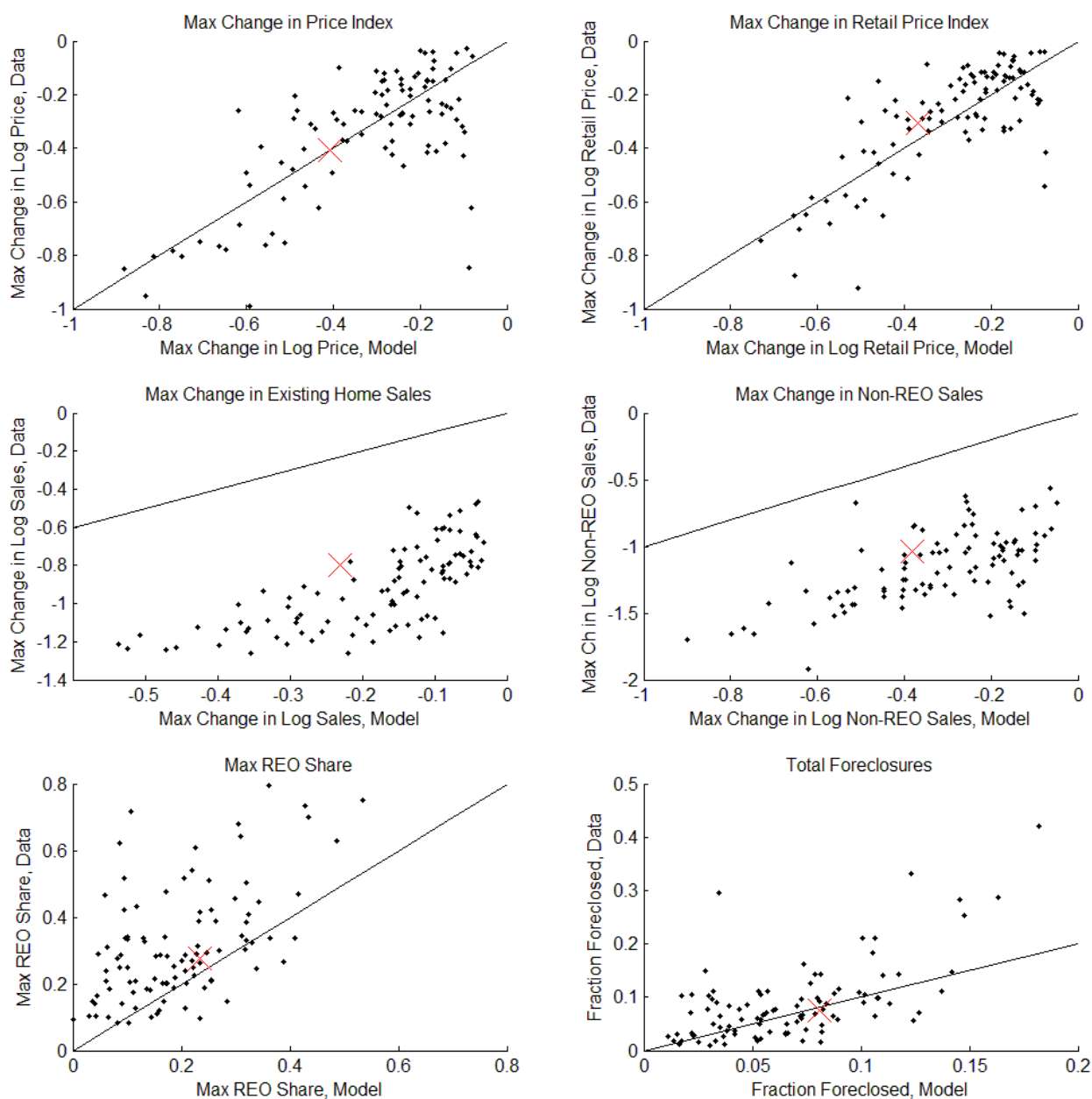
$$B(t) = \beta B(t+1) + b + q_b(\mu(t)) \frac{1-\theta}{\lambda[1-\beta(1-\gamma)]} e^{-\lambda(h_n(t)-a)}. \quad (\text{B.2.2})$$

Again, we leave all other equations unchanged.

3. Finally, we run an experiment in which we include only compositional effects. For this experiment, we shut down both the market tightness and choosey buyer effects by modifying the law of motion for the stock of buyers and the Bellman equation for the buyer's value function in the manner described above.

B.2.2 Cross-Markets Analysis With 10% REO Discount

Figure B.2.1 shows the results of the same calibration procedure in 2.7 for a 10% REO discount instead of a 20% REO discount. The lower REO discount weakens the compositional effect whereby a large REO share reduces the aggregate price index by mechanically placing more weight on properties that sell at a discount. The lower discount also weakens the choosey buyer effect, since the benefit of waiting for a foreclosure is reduced somewhat (though it still grows substantially in the downturn). Consequently, to



Note: Scatter plots of data vs. simulation results for 97 MSAs in regression analysis for a 10 percent discount. The red X represents the national simulation and each black dot is an MSA. The 45-degree line illustrates a perfect match between the model and the data. The variable being plotted shown in each plot's title. Data is fully described in appendix B.4. The calibration methodology described in text and appendix B.4.

Figure B.2.1: Cross-MSA Simulations vs. Data: 10% REO Discount

Table B.3.1: Judicial vs. Non-Judicial States

Dependent Variable	$\Delta \log (P)$	$\Delta \log (P_{\text{Retail}})$
Data	0.084 (0.035)**	0.053 (0.025)**
Model With Backlogs	0.014 (0.005)***	0.002 (0.002)
Model No Backlogs	0.008 (0.004)*	0.004 (0.002)*
N	45	45

Notes: * = 10% Significance, ** = 5% Significance *** = 1% significance. All standard errors are robust to heteroskedasticity. Every reported coefficient is for the judicial state dummy in a regression that includes a linear and quadratic term for change in log price 2003-2006, z score for share with LTV over 80 percent and its interaction with change in log price 2003-2006, and z score for share with second mortgage and its interaction with change in log price 2003-2006. These regressions do not include the Saiz (2010) variables, which are not available at the state level. The columns differ by dependent variables. The rows differ by data source: the first row shows the actual CoreLogic data, the second row uses simulated dependent variable data from a model in which judicial states have a backlog, and the third row uses simulated dependent variable data from a model in which judicial states have no backlog. Every regression has 45 states as described in Appendix B.4. We use data from Mian et al. (2012), which they obtained from RealtyTrac.com, to categorize states as judicial foreclosure or non-judicial foreclosure states. Using this methodology, Connecticut, Delaware, Florida, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Nebraska, New Jersey, New York, North Dakota, Ohio, Pennsylvania, South Carolina, and Vermont are judicial states.

match the non-linearity in price declines relative to the size of the preceding boom, the non-compositional effects of foreclosure – the market tightness effect – must be larger. This results in a longer average time out of the market of 1.3 years relative to 1.05 years for a 20% steady state discount. Additionally, the permanent price decline is 22.4%, which is slightly bigger than the 21.5% for a 20% discount. Foreclosures thus exacerbate the downturn by 50%. The permanent price decline in the model remains high in order to fit the non-linearity shown in Figure 2.2.2.

Interestingly, because the compositional effect is reduced, the extent to which foreclosures exacerbate the retail price decline is increased from 28.7% to 37.5%. Intuitively, with the compositional effect weakened the retail price declines must be stronger. Unfortunately, this results in a worse average fit for retail price declines, as shown in panel B of Figure B.2.2. Thus while the model fit for overall price declines is roughly comparable to the 20% steady state price decline case, we prefer the 20% case shown in the main text.

B.3 Judicial vs. Non-Judicial States

Table B.3.1 shows the coefficient on judicial state of running a regression similar to equation (2.2.1) with a judicial state dummy as described in the table note.¹ The first row shows the actual data, while the

¹As with the main text, the results are largely unchanged if we used weighted least squares and weight by the owner-occupied housing stock.

second row shows the results of a model with a backlog of $\phi = 3,400$ for judicial states and no backlog for non-judicial states and the third row shows the results of a model with no backlog for judicial or non-judicial states. Adding backlogs to the model is a step in the right direction, but the model is still an order of magnitude short of the data.

As a result, we speculate that this implies that backlogs cannot be the whole story in judicial states – there must be some reduction in the incidence of foreclosures as banks respond to the long foreclosure timelines. While it is possible that $\phi > 3,400$ in judicial states – something we cannot simulate because of numerical issues – the results of higher backlogs for a national calibration shown in Figure 2.8.2 suggests that even with a much narrower foreclosure pipeline it is not possible to get that judicial states have a log price decline that is .084 smaller from foreclosure backlogs alone.

B.4 Data Sources and Calculations

Data

The main data source is proprietary data from CoreLogic, which we supplement with data from the U.S. Census, Saiz (2010), and Mian et al. (2012).

CoreLogic provides us with a monthly data set for the nation, 50 states, and the 100 largest MSAs² that for 2000-2011 includes:

²By CBSA code and name, they are: 10420 Akron, OH; 10580 Albany-Schenectady-Troy, NY; 10740 Albuquerque, NM; 10900 Allentown-Bethlehem-Easton, PA-NJ; 12060 Atlanta-Sandy Springs-Marietta, GA; 12420 Austin-Round Rock-San Marcos, TX; 12540 Bakersfield-Delano, CA; 12580 Baltimore-Towson, MD; 12940 Baton Rouge, LA; 13644 Bethesda-Rockville-Frederick, MD; 13820 Birmingham-Hoover, AL; 14484 Boston-Quincy, MA; 14860 Bridgeport-Stamford-Norwalk, CT; 15380 Buffalo-Niagara Falls, NY; 15764 Cambridge-Newton-Framingham, MA; 15804 Camden, NJ; 16700 Charleston-North Charleston-Summerville, SC; 16740 Charlotte-Gastonia-Rock Hill, NC-SC; 16974 Chicago-Joliet-Naperville, IL; 17140 Cincinnati-Middletown, OH-KY-IN; 17460 Cleveland-Elyria-Mentor, OH; 17820 Colorado Springs, CO; 17900 Columbia, SC; 18140 Columbus, OH; 19124 Dallas-Plano-Irving, TX; 19380 Dayton, OH; 19740 Denver-Aurora-Broomfield, CO; 19804 Detroit-Livonia-Dearborn, MI; 20764 Edison-New Brunswick, NJ; 21340 El Paso, TX; 22744 Fort Lauderdale-Pompano; Beach-Deerfield Beach, FL; 23104 Fort Worth-Arlington, TX; 23420 Fresno, CA; 23844 Gary, IN; 24340 Grand Rapids-Wyoming, MI; 24660 Greensboro-High Point, NC; 24860 Greenville-Mauldin-Easley, SC; 25540 Hartford-West Hartford-East Hartford, CT; 26180 Honolulu, HI; 26420 Houston-Sugar Land-Baytown, TX; 26900 Indianapolis-Carmel, IN; 27260 Jacksonville, FL; 28140 Kansas City, MO-KS; 28940 Knoxville, TN; 29404 Lake County-Kenosha County, IL-WI; 29820 Las Vegas-Paradise, NV; 30780 Little Rock-North Little Rock-Conway, AR; 31084 Los Angeles-Long Beach-Glendale, CA; 31140 Louisville-Jefferson County, KY-IN; 32580 McAllen-Edinburg-Mission, TX; 32820 Memphis, TN-MS-AR; 33124 Miami-Miami Beach-Kendall, FL; 33340 Milwaukee-Waukesha-West Allis, WI; 33460 Minneapolis-St. Paul-Bloomington, MN-WI; 34980 Nashville-Davidson-Murfreesboro-Franklin, TN; 35004 Nassau-Suffolk, NY; 35084 Newark-Union, NJ-PA; 35300 New Haven-Milford, CT; 35380 New Orleans-Metairie-Kenner, LA; 35644 New York-White Plains-Wayne, NY-NJ; 35840 North Port-Bradenton-Sarasota, FL; 36084 Oakland-Fremont-Hayward, CA; 36420 Oklahoma City, OK; 36540 Omaha-Council Bluffs, NE-IA; 36740 Orlando-Kissimmee-Sanford, FL; 37100 Oxnard-Thousand Oaks-Ventura, CA; 37764 Peabody, MA; 37964 Philadelphia, PA; 38060 Phoenix-Mesa-Glendale, AZ; 38300 Pittsburgh, PA; 38900 Portland-Vancouver-Hillsboro, OR-WA; 39100 Poughkeepsie-Newburgh-Middletown, NY; 39300 Providence-New Bedford-Fall River, RI-MA; 39580 Raleigh-Cary, NC; 40060 Richmond, VA; 40140 Riverside-San Bernardino-Ontario, CA; 40380 Rochester, NY; 40900 Sacramento-Arden-Arcade-Roseville, CA; 41180 St. Louis, MO-IL; 41620 Salt Lake City, UT; 41700 San Antonio-New Braunfels, TX; 41740 San Diego-Carlsbad-San Marcos, CA; 41884 San Francisco-San Mateo-Redwood City, CA; 41940 San Jose-Sunnyvale-Santa Clara, CA; 42044 Santa Ana-Anaheim-Irvine, CA; 42644 Seattle-Bellevue-Everett, WA; 44140 Springfield, MA; 44700 Stockton, CA; 45060 Syracuse, NY; 45104 Tacoma, WA; 45300 Tampa-St. Petersburg-Clearwater, FL; 45780 Toledo, OH; 46060 Tucson, AZ; 46140 Tulsa, OK; 47260 Virginia Beach-Norfolk-Newport News, VA-NC; 47644 Warren-Troy-Farmington Hills, MI; 47894 Washington-Arlington-Alexandria, DC-VA-MD-WV; 48424 West Palm Beach-Boca Raton-Boynton Beach, FL; 48864 Wilmington, DE-MD-NJ; 49340 Worcester, MA.

- The CoreLogic home price index and non-distressed home price index estimated from public records. We refer to these as the aggregate and retail price indices. The CoreLogic non-distressed price index differs slightly from the retail price index in the model because it excludes short sales, which we count as non-REO sales.
- The number of pre-foreclosure filings and completed foreclosure auctions estimated from public records.
- Sales counts for REOs, new houses, non-REO and non-short sale resales, and short sales estimated from public records. Because short sales are not reported separately for much of the time frame represented by the data, we combine short sales and resales into a non-REO existing home sales measure which we call retail sales. We calculate existing home sales by adding REO and retail sales. We also use this data to construct the REO share of existing home volume, which we seasonally adjust.
- Estimates of 7 quantiles of the combined loan-to-value distribution for active mortgages: under 50%, 50%-60%, 60%-70%, 70%-80%, 80%-90%, 90%-100%, 100%-110%, and over 110%. These statistics are compiled by CoreLogic using public records and CoreLogic’s valuation models.
- First lien originations and first lien refinancings estimated using public records.
- Over-90-day-delinquent loans, loans in foreclosure, and active loans estimated using a mortgage-level database. We use the raw counts to construct the share of active loans that are over 90 days delinquent and in foreclosure.
- The mean number of days on the market for listed homes and closed sales estimated using Multiple Listing Service data.

We seasonally adjust the raw CoreLogic house price indices, foreclosure counts, sales counts, and delinquent and in-foreclosure loan shares using the Census Bureau’s X-12 ARIMA software with an additive seasonal factor. For the state and county-level sales counts, auctions counts, days on the market, and REO share, we smooth the data using a 5 month moving average (2 months prior, the current month, and 2 months post) to remove any blips in the data caused by irregular reporting at the county level.

For the calibration of the loan balance distribution and initial number of mortgages with high LTV ratios, we adjust the CoreLogic data using data from the American Community Survey as tabulated by the Census. The CoreLogic data only covers all active loans, while our model corresponds to the entire owner-occupied housing stock. Consequently, we use the ACS 3-year 2005-2007 estimates of the owner-occupied housing stock and fraction of houses with a mortgage at the national, state, and county level, which we aggregate

to the MSA level using MSA definitions.³ From this data, we construct the fraction of owner-occupied housing units with a mortgage and the fraction of owner-occupied housing units with a second lien or home equity loan. We use these estimates to adjust the loan balance distribution so it represents the entire owner-occupied housing stock and in our regressions to construct the fraction of owner-occupied houses with over 80% LTV.

The LTV data is first available for March 2006, which roughly corresponds to the eve of the housing bust as the seasonally-adjusted national house price index reached its peak in March 2006. To approximate the size of the bubble, we average the seasonally-adjusted price index for March-May 2001, March-May 2003, March-May 2006, and March-May 2011 to calculate the change in log prices for 2001 to 2006, 2003 to 2006, and 2006 to 2011. We use these variables in our regressions and to estimate the relative size of the shock for each geographical area.

We also estimate the maximum log change in seasonally-adjusted prices, smoothed and seasonally-adjusted volume, and seasonally-adjusted time to sale as well as the maximum REO share for each geographical area. We estimate the minimum value between March 2006 and December 2011 and the maximum value between January 2002 and December 2007. We implement these restrictions so that the addition of counties to the CoreLogic data set prior to 2002 does not distort our results. We calculate the fraction of the owner-occupied housing stock that was foreclosed upon by adding up completed foreclosure auctions between March 2006 and December 2012 and dividing by the owner-occupied housing stock in 2006 as calculated from the ACS adjusted for CoreLogic’s approximately 85% coverage, which is assumed to be constant across locations. Again, our results are not sensitive to the choice of dates.

From the 100 MSAs and 50 states, we drop two MSAs and five states. The Birmingham, Alabama MSA is dropped because a major county stopped reporting to CoreLogic in the middle of the downturn, and the Syracuse New York MSA is dropped because loan balance distribution data is not available for this MSA in 2006. Maine, Vermont, and South Dakota are dropped because loan balance distribution data is not available for these states in 2006. For the cross-state analysis, we focus on the continental U.S. and omit Alaska and Hawaii.

We finally merge data from Saiz (2010) into the MSA data. The Saiz data includes his estimate of unusable land due to terrain, the housing supply elasticity, and the Wharton Land-Use Regulation Survey score for each MSA. We are able to match every MSA we have data on except for Sacramento and Honolulu.

Loan Balance Distribution Calibration

³The 3-year ACS estimates include estimates of the housing stock and houses with a mortgage for all counties with over 20,000 residents. For a few MSAs, one or more small counties are not included in the ACS data. The bias on our constructed estimates of the fraction of owner-occupied homes with a mortgage and with a second lien or home equity loan due to these small missing counties is minimal.

We use a minimum-distance methodology to calibrate the loan balance distribution for each geography. From the 7 quantiles given to us by CoreLogic and the Census Data on the number of owner-occupied homes without a mortgage, we construct a CDF of 6 points: the fraction of loans with under 50% LTV, under 60% LTV, under 70% LTV, under 80% LTV, and under 100% LTV. We then construct a norm for the distance between the Beta distribution and the empirical CDF. Because the upper tail of the distribution is most critical for our amplification channel, we weight the under 50%, under 60%, and under 70% parts of the distribution by .1 and the under 80%, under 90%, and under 100% by .2. We then choose b_a and b_b , the parameters of the Beta distribution, to minimize this norm. The resulting fit is close enough that our results are robust to alternate weightings of the norm.

Sources of Calculations in the Text

All figures in the introduction are tabulated from the CoreLogic data as described above.

For the calibration of the housing market model, the median tenure for owner occupants of approximately 9 years comes from table 3-9 of the American Housing Survey reports for 1997-2005. The 20% REO discount comes from Campbell et al.'s (2011) online appendix. They report an average discount over 1987-2009 of 26%. In table A6, they estimate this by year and show that in current housing cycle it was as low as 22.6% in 2005 and as high as 35.4% in 2009. 20 percent is thus a reasonable discount.

To determine γ_I , the incidence of income shocks for houses in negative equity, we divide the seasonally adjusted number of foreclosures by the maximum seasonally adjusted number of homes in negative equity in the CoreLogic data. The mean annual incidence is $\gamma_I = 8.6\%$.

To get that interest rates decrease the hazard of default for under-water borrowers from 8.6 percent to 7.1 percent, we use data from Bajari et al. (2010) combined with standard mortgage amortization schedules. We begin by assuming that all mortgages are at 7 percent interest rates and will be refinanced to 4 percent. The average mortgage is somewhat below 7 percent, but we choose 7 percent to reflect that some ARMs reset at quite high rates and because we want to simulate the largest possible impact of a refinancing. Assuming houses are bought with 20 percent down in steady state, a 7 percent mortgage has a monthly payment of \$1,217.50 while a 4 percent mortgage has a monthly payment of \$873.67 according to standard amortization schedules and formulas. Bajari et al. report that the mean loan in their data set has a payment-to-income ratio of .312. Assuming the \$1,217.50 monthly payment matches this ratio, monthly income is \$3,902.24. A reduction of the monthly payment to \$873.67 reduces the payment-to-income ratio by .088. Bajari et al. estimate that a one standard deviation change in the payment-to-income ratio – equivalent to a .124 change – reduces the hazard of default by 17.5 percent. Assuming linearity a reduction of .088 will reduce the hazard of default by 12.435 percent. Given the initial hazard of 8.6 percent, this implies a default hazard with a reduced interest rate of 7.1 percent.

For the principal reductions, we assume a \$100 billion dollar principal reduction. 21.5 million households potentially under water implies an approximately \$5,000 principal reduction for each house. If all 50 million households with a mortgage received the reduction, this would be only a \$2,000 principal reduction for each house.

Appendix C

Supplement to Chapter 3

C.1 No Nash Equilibria with Asymmetric Behavior within Countries

In the text, we assumed that all firms in country j behaved in the same way ($q_{ij} = q_j$ and $k_{ij} = k_j$). This led us to analyze conditions (3.2.5) and (3.2.6). In this appendix, we prove that there can be no Nash equilibria where firms within any country j behave differently. To do this, we must analyze conditions (3.2.3) and (3.2.4). We first establish through three lemmas that in any nontrivial Nash equilibrium¹ every firm in every country is active ($q_{ij} > 0$) and produces a quality above the minimum level ($k_{ij} > 0$). Since the optimum for every country is interior, conditions (3.2.3) and (3.2.4) simplify and we prove that their unique solution has all firms within a country producing the same quantity at the same quality.

Lemma C.1.1 *There can be no non-trivial equilibrium in which a firm produces a positive quantity of the minimum quality ($q_{ij} > 0$ implies $k_{ij} > 0$).*

Proof. If it is optimal to produce at minimum quality ($k_{ij} = 0$), then since $c'(0) = 0$ condition (3.2.3) requires that $\theta \frac{q_{ij}}{Q^j} \leq 0$. But each of the factors to the left of this inequality is strictly positive since, by hypothesis, $q_{ij} > 0$. So the inequality can never hold. Therefore, every firm with $q_{ij} > 0$ must have $k_{ij} > 0$.

■

Intuitively, since the cost function is flat at the origin but inverse demand is strictly increasing in quality when the price is non-zero, an active firm producing a minimal quality can always increase his profit by

¹Conditions (3.2.3) and (3.2.4) (and, by extension, conditions (3.2.5) and (3.2.6)) hold only in non-trivial equilibria where inverse demand is not the zero function.

marginally increasing his quality choice. At the margin, costs will remain the same but the price will increase.

Lemma C.1.2 *There can be no non-trivial equilibrium in which an active country can have an inactive firm ($Q^j > 0$ implies $q_{ij} > 0$ at every firm i in country j).*

Proof. Suppose that $q_{ij} = 0$ and $Q^j > 0$. In that case, one or more of the rival firms is producing a strictly positive amount. Label as firm i' the active firm with the smallest quality. Hence, $k_{i'j} - R^j \leq 0$. Since firm i' produces a strictly positive amount, its first-order condition in (3.2.3) must hold with equality. Since the terms $q_{i'j}U'(Q)$ and $q_{i'j}\theta(k_{i'j} - R^j)/Q$ are respectively strictly and weakly negative, (3.2.3) implies $a + \theta R^j - U(Q) - c(k_{i'j}) > 0$. But then firm i could produce a marginal amount at quality $k_{i'j}$ and make positive profits. This a profitable deviation. ■

Therefore, we see that active countries cannot have inactive firms. Similar to Theorem 3.2.2, we may also show that there cannot exist Nash equilibria with inactive countries.

Lemma C.1.3 *There exist no non-trivial pure strategy Nash equilibria in which some countries produce no output ($Q^j > 0$ for every country j).*

Proof. Suppose that $Q^j = 0$. Consider an active country h with collective reputation R^h . If all firms within country h produce the same quality, then Theorem 3.2.2 applies. Suppose firms in country h produce asymmetric qualities. Then there exists a firm i in country h such that $k_{ih} > R^h$. By assumption of Nash equilibrium, the unit margin of firm i must be nonnegative. Since the cost function is strictly increasing, this implies that $a + \theta R^h - U(Q) - c(R^h) > 0$. But then a firm in country j could produce a marginal amount at quality R^h and make positive profits. This is a profitable deviation. ■

Given these results, we can show there exist no asymmetric Nash equilibria.

Theorem C.1.4 *There exist no non-trivial pure strategy Nash equilibria in which firms within the same country produce different qualities and/or outputs.*

Proof. If there is such an equilibrium, then by the three previous lemmas, the first-order conditions of each firm must hold with equality. That is, the following equations must hold in equilibrium for every country j :

$$\Gamma^j(k_{ij}; Q, Q^j, R^j) \equiv [k_{ij} - (R^j + \frac{Q^j U'(Q)}{\theta})]c'(k_{ij}) + \{a + \theta R^j - U(Q) - c(k_{ij})\} = 0. \quad (\text{C.1.1})$$

We will first show that it is not possible in equilibrium for any two firms in the same country j facing the same Q, Q^j , and R^j to have different qualities. The term in braces is the unit margin. It must be nonnegative

or the profit in (2) would be strictly negative and the firm could deviate profitably to zero production. This implies that k_{ij} must satisfy:

$$0 < k_{ij} \leq R^j + \frac{Q^j U'(Q)}{\theta}, \quad (\text{C.1.2})$$

where the strict inequality is a consequence of our three lemmas and the weak inequality follows since the cost function is strictly increasing ($c'(k_{ij}) > 0$ for $k_{ij} > 0$). In equilibrium, every firm ($i = 1, \dots, n_j$) in country j will have $\Gamma^j(k_{ij}; Q, R^j) = 0$. However, this equation cannot have more than one root over the relevant range. We see:

$$\frac{\partial \Gamma^j}{\partial k_{ij}}(k_{ij}; Q, Q^j, R^j) = [k_{ij} - (R^j + \frac{Q^j U'(Q)}{\theta})] c''(k_{ij}). \quad (\text{C.1.3})$$

This indicates $\Gamma^j(k_{ij}; Q, Q^j, R^j)$ is a continuous function and strictly decreasing over the interval $[0, R^j + \frac{Q^j U'(Q)}{\theta}]$. The claim therefore follows by Rolle's theorem.

Hence, there can be no more than one root. So, for any given equilibrium with its (Q, R^j) a unique k_{ij} satisfies equation (C.1.1). But since equation (C.1.1) must hold for *every* firm in country j , each firm must choose the same quality in this equilibrium. Denote it $k_j(Q, R^j)$. Moreover, since every firm will be active and reputed quality will equal $R^j = k_j(Q, R^j)$, equation (3.2.3) implies that:

$$q_{ij} = \frac{a + \theta R^j - U(Q) - c(k_j(Q, R^j))}{U'(Q)}. \quad (\text{C.1.4})$$

Since the right-hand side of this equation is independent of i , every firm will produce the same quantity in this equilibrium. Denote it $q_j(Q, R^j)$. ■

Appendix D

Supplement to Chapter 4

D.1 Existence and Uniqueness of a Non-trivial Symmetric Solution of the First-Order Conditions

In this appendix, we establish that there exists a unique solution to equations (4.2.4) and (4.2.5) in which firms within a sector follow symmetric strategies and receive a strictly positive price for the product.

Theorem D.1.1 *There exists a non-trivial solution to equations (4.2.4) and (4.2.5) in which quality and output choices within sectors are symmetric.¹ This solution is unique.*

Proof. Given the symmetry we are considering, we can write the Kuhn-Tucker conditions as

$$q_j \geq 0, \quad P^j(Q, k_j) + q_j P_1^j(Q, k_j) - c(k_j) \leq 0, \quad \text{c.s.} \quad (\text{D.1.1})$$

$$Q = \sum_{j=1}^N n_j q_j \quad (\text{D.1.2})$$

and

$$k_j \geq 0, \quad q_j [P_2^j(Q, k_j) - n_j c'(k_j)] \leq 0, \quad \text{c.s.} \quad (\text{D.1.3})$$

where $R^j = k_j$.

To begin, consider the solution to equation (D.1.3) given a particular Q and q_j . If the firm is inactive ($q_j = 0$), then any quality choice will solve the equation. Now let $\tilde{k}_j = (c')^{-1}(\theta/n_j)$. If the firm is active, it is clear that for all $0 \leq Q < \bar{Q}(\tilde{k}_j)$, \tilde{k}_j is the unique nonzero solution to equation (D.1.3). For $Q \geq \bar{Q}(\tilde{k}_j)$,

¹Trivial equilibria are those in which firms receive a price of zero for the product.

there is no nonzero solution.

Let us now consider the system of equations given by (D.1.1) and (D.1.2) when $k_j = \tilde{k}_j$ for all j . Note that $P^j(0, \tilde{k}_j) - c(\tilde{k}_j) = \theta \tilde{k}_j - c(\tilde{k}_j) > 0$ for all j .² Next, define $\hat{Q}(\tilde{k}_j) < \bar{Q}(\tilde{k}_j)$ such that $P(\hat{Q}(\tilde{k}_j), \tilde{k}_j) = c(\tilde{k}_j)$. Then the solution to (D.1.1) is $q_j = 0$ for all $Q \geq \hat{Q}(\tilde{k}_j)$. By totally differentiating equation (D.1.1) with respect to Q , we find that for $Q < \hat{Q}(\tilde{k}_j)$

$$\frac{dq_j}{dQ} = -\frac{P_1^j(Q, \tilde{k}_j) + q_j P_{11}^j(Q, \tilde{k}_j)}{P_1^j(Q, \tilde{k}_j)} = -\frac{G'(Q) + q_j G''(Q)}{G'(Q)} < 0 \quad (\text{D.1.4})$$

Now let $f(Q) = \sum_{j=F,I} n_j q_j(Q)$ and $Q^H = \max\{\hat{Q}(\tilde{k}_F), \hat{Q}(\tilde{k}_I)\}$. Then $f(Q)$ is continuous and strictly decreases from $f(0) > 0$ to $f(Q^H) = 0$ and $f(Q) = 0$ for all $Q \geq Q^H$. Thus, the curve will cross the 45° line exactly once at some $0 < Q^* < Q^H$. Let $k_j^* = \tilde{k}_j$ and $q_j^* = q_j(Q^*)$ for all j . If $Q^* < \hat{Q}(\tilde{k}_j) < \bar{Q}(\tilde{k}_j)$, then $q_j^* > 0$, and if $Q^* \geq \hat{Q}(\tilde{k}_j)$, then $q_j^* = 0$. The profile $((k_j^*)_{j=F,I}, (q_j^*)_{j=F,I}, Q^*)$ satisfies the Kuhn-Tucker necessary conditions. Note that at least one market will be active, since $Q^* < Q^H$. ■

D.2 Nonexistence of Asymmetric Nash Equilibria

In this appendix, we show that the symmetric outcome is the only possible Nash equilibrium.

Theorem D.2.1 *There exist no nontrivial pure strategy Nash equilibria in which firms in active markets produce different qualities and/or outputs.*

The proof first proceeds by establishing two lemmas which demonstrate that in any nontrivial equilibrium, (1) active firms never produce the minimum quality and (2) active sectors cannot have inactive firms.

Lemma D.2.2 *There can be no nontrivial equilibrium (symmetric or otherwise) in which a firm produces a positive quantity of the minimum quality.*

Proof. If it is optimal to produce at minimum quality ($k_{ij} = 0$), then since $c'(0) = 0$, condition (4.2.5) requires that $P_2(Q, R^j) \frac{q_{ij}}{Q^j} \leq 0$. But each of the factors to the left of this inequality is strictly positive since, by hypothesis, $q_{ij} > 0$ and we are considering nontrivial equilibria. So the inequality can never hold. Therefore, every firm with $q_{ij} > 0$ must have $k_{ij} > 0$. ■

Intuitively, because the cost function is flat at the origin but inverse demand is strictly increasing in quality when the price is nonzero, an active firm producing a minimal quality can always increase its profit

²Let $\hat{k} = (c')^{-1}(\theta)$. Then since $c(0) = 0$ and the cost function is convex, we know $k - c(k)$ is maximized at \hat{k} . Since $\tilde{k}_j \leq \hat{k}$ for all j , it follows that $\theta \tilde{k}_j - c(\tilde{k}_j) > 0$ for all j .

by marginally increasing its quality choice. At the margin, costs will remain the same but the price will increase.

Lemma D.2.3 *There can be no nontrivial equilibrium (symmetric or otherwise) in which an active market can have an inactive firm.*

Proof. Suppose that $q_{ij} = 0$ and $Q^j > 0$. In that case, one or more of the rival firms is producing a strictly positive amount. Label as firm i' the active firm with the smallest quality. Hence, $k_{i'j} - R^j \leq 0$. Because firm i' produces a strictly positive amount, its first-order condition in (4.2.4) must hold with equality. Because the terms $q_{i'j}P_1^j(Q, R^j)$ and $q_{i'j}P_2^j(Q, R^j)(k_{i'j} - R^j)/Q$ are respectively strictly and weakly negative, (4.2.4) implies $P^j(Q, R^j) - c(k_{i'j}) > 0$. But because the cost function is strictly increasing, $P(Q, R^j) - c(0) > 0$, and this same complementary slackness condition, which must hold for firm i as well, implies that $q_{ij} > 0$, contradicting the hypothesis that $q_{ij} = 0$. ■

Now consider an active sector j . By the two previous lemmas, the first-order conditions of each firm in this sector must hold with equality. That is, the following equation must hold in equilibrium:

$$[k_{ij} - (R^j - \frac{QG'(Q)}{\theta R^j})]c'(k_{ij}) + \{\theta R^j - G(Q) - c(k_{ij})\} = 0 \quad (\text{D.2.1})$$

Define the left-hand side as $\Gamma^j(k_{ij}; Q, R^j)$. In equilibrium, every firm ($i = 1, \dots, n_j$) will have $\Gamma^j(k_{ij}; Q, R^j) = 0$. However, this equation cannot have more than one root. We see that $\frac{\partial \Gamma^j}{\partial k_{ij}}(k_{ij}; Q, R^j) = [k_{ij} - (R^j - \frac{QG'(Q)}{\theta R^j})]c''(k_{ij})$, and (D.2.1) requires that at any root, the first factor in $\frac{\partial \Gamma^j}{\partial k_{ij}}(k_{ij}; Q, R^j)$ must be strictly negative.³

Hence, there can be no more than one root. So for any given equilibrium with its (Q, R^j) , a unique k_{ij} satisfies equation (D.2.1). But because equation (D.2.1) must hold for *every* firm, each firm must choose the same quality in this equilibrium. Denote it $k_j(Q, R^j)$. Moreover, as every firm will be active and reputed quality will equal $R^j = k_j(Q, R^j)$, equation (4.2.4) implies that $q_{ij} = \frac{\theta R^j - G(Q) - c(k_j(Q, R^j))}{-G'(Q)}$. Because the right-hand side of this equation is independent of i , every firm will produce the same quantity in this equilibrium. Denote it $q_j(Q, R^j)$. This completes the proof of the theorem.

³The term in braces in (D.2.1) must be strictly positive, since $q_{ij} > 0$; condition (D.1.1) therefore holds with equality, and $P_1 < 0$ by assumption.